Mars Express
Mapping With OMEGA

125
YEARS OF GLOBAL
Science

AAAS

Science

SPECIAL ISSUE
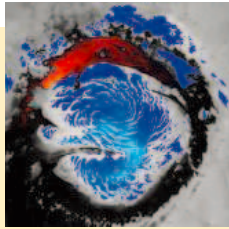
MARS EXPRESS: OMEGA

False-color image of the north polar region of Mars in summer, showing its composition as inferred by the OMEGA/Mars Express visible and near-infrared imager. The cap is made of water ice (blue) mixed with mineral grains (shades of gray), with dark zones of ice-free minerals within which vast areas of gypsum (red), a hydrated sulfate, have been discovered. [Image: © Institut d'Astrophysique Spatiale]

1540

1548

Science

1570

1572 &
1625

Science

REPORTS CONTINUED

1612

1618

Science
www.scienceonline.org

sciencenow www.sciencenow.org DAILY NEWS COVERAGE

The Great Mountain Builder
Some mountain ranges owe their height to very cold climates.

A Hot Stellar Womb
Strong magnetic fields may play a role in the earliest formation of stars.

Outdoing Mother Nature
Humans erode more earth than all natural processes combined.

science's next wave www.nextwave.org CAREER RESOURCES FOR YOUNG SCIENTISTS

Countering nonconventional threats.

GLOBAL: **Next Wave Special Issue—Science Careers in National Security** *Edited by A. Kotok*
Next Wave explores careers in the science of today's threats to national security.

GLOBAL: **Opportunities for Scientists at the U.S. Defense Intelligence Agency** *J. Kling*
The DIA offers American scientists a way to contribute to the defense of the U.S. and its allies.

GLOBAL/CANADA: **Science in Defense—Canadian Careers in National Security Research** *A. Fazekas*
Several Canadian agencies employ scientists to ensure the safety of its national borders.

GLOBAL/UK: **A Scientist as a Knowledge Agent** *A. Forde*
A former geosciences researcher now works for the UK's Defence Science and Technology Laboratory.

MISCINET: **The Beauty of Statistics** *E. Francisco*
Francisco Samaniego hopes to encourage more minority students to consider careers in statistics.

CAREER DEVELOPMENT CENTER: **Small-College Shenanigans** *GrantDoctor*
For early-career scientists at small colleges, the biggest barrier to research productivity can be the institution itself.

science's sage ke www.sageke.org SCIENCE OF AGING KNOWLEDGE ENVIRONMENT

GENETICALLY ALTERED MICE: **Park2^{tm1Rpa} Mice** *J. Fuller*
Mice designed to model a heritable form of Parkinson's disease do not exhibit parkinsonism.

NEWS FOCUS: **Sugar Rush** *M. Leslie*
Potential life-extending enzyme cranks up glucose synthesis.

NEWS FOCUS: **Early Warning** *R. J. Davenport*
Memory fades when β amyloid accumulates inside neurons.
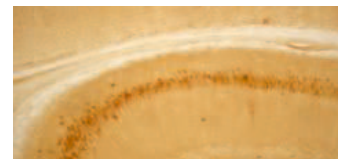
The inside track on Alzheimer's disease.

science's stke www.stke.org SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

Crystal structure of calcium channel subunits.

PERSPECTIVE: **Alone at Last! New Functions for Ca^{2+} Channel β Subunits?** *M. Rousset, T. Cens, P. Charnet*
β subunits exhibit regulatory activities that are independent of the pore-forming α subunit.

PROTOCOL: **Utilizing the Split-Ubiquitin Membrane Yeast Two-Hybrid System to Identify Protein-Protein Interactions of Integral Membrane Proteins** *K. Iyer, L. Bürkle, D. Auerbach, S. Thaminy, M. Dinkel, K. Engels, I. Stagljar*
Reconstitution of ubiquitin allows screening for membrane protein–binding partners.

Separate individual or institutional subscriptions to these products may be required for full-text access.

Thin But Tough Networks

Additives that can form thin protective films on surfaces are typically added to lubricants in order to reduce the wear between moving parts. For steel engines, the primary ones are the zinc phosphates, but their breakdown by-products poison catalytic converters and they do not work well in aluminum engines. Using simulations, **Mosey et al.** (p. 1612) show that at the high pressures that occur during a compressing cycle in the engine, the zinc changes coordination number and forms chemically connected networks. Their results explain why other divalent cations, such as calcium, cannot be substituted for zinc, and why these additives do not work well in aluminum engines, where the strength of the alloys is such that the pressures do not get high enough to form the antiwear films.

Prolonging Antigen Presentation

It has been assumed that antigen-presenting cells must have exceptionally well developed capacities for proteolysis because they must degrade protein antigens to perform their function. However, **Delamarre et al.** (p. 1630) now find that the most efficient of the antigen-presenting cells (dendritic cells and B cells) harbor exceptionally low concentrations of lysosomal proteases when these levels are compared to those of macrophages. Dendritic cells also contain endogenous protease inhibitors that further attenuate their proteolytic potential. Remarkably, the levels of other lysosomal hydrolases in dendritic cells are similar to those found in macrophages. Thus, whereas macrophages rapidly degrade the antigens they encounter, dendritic cells may protect the very same antigens, facilitating their dissemination to and survival in secondary lymphoid organs.

Sex and Smell

In the antennae of the insect olfactory system, there exist two distinct chemical perception mechanisms. The so-called "generalist" system recognizes odorants from foods and plants and is made up of the olfactory receptor family with many different genes. The second perception mechanism, the "specialist" system, detects pheromones from insects of the same species. **Nakagawa et al.** (p. 1638, published online 3 February 2005) report that in the silk moth, coexpression of pheromone receptors with a receptor from the generalist insect olfactory receptor
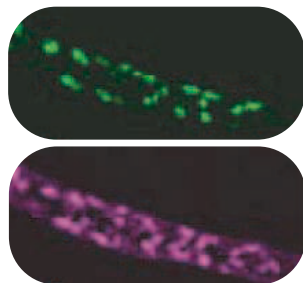
subfamily promotes the functional expression of pheromone receptors and confers ligand-stimulated nonselective cation channel activity.

Domesticating Pigs Seven Times Over

DNA sequencing has revolutionized the study of the domestication patterns of animals and plants by humans. Archaeological evidence suggests that domestication of wild boar took place principally in Asia. **Larson et al.** (p. 1618) focus on the origins and spread of the domesticated pig by examining mitochondrial DNA sequences from 687 wild, feral, and domestic pigs (across the entire natural range of wild boar) and combining these data with phylogenetic analyses. Pig domestication took place at least seven times in areas across Eurasia, including in previously unknown centers in India, Burma-Thailand, Central Italy, and Wallacea–New Guinea.

Mapping Magnetic Galaxies

Many galaxies in the universe show signs of complex magnetic structures that are difficult to measure and are not well understood. One way to map out the magnetism is by means of the Faraday effect, in which the plane of polarization in an electromagnetic wave is rotated by a magnetic field. **Gaensler et al.** (p. 1610) report their measurement of polarized radio emissions from distant sources behind the Large Magellanic Cloud (LMC). The survey of 291 radio sources showed that the LMC has an axisymmetric spiral magnetic field that exhibits noticeable fluctuations. This analysis suggests that the field is produced by a cosmic-ray–driven dynamo mechanism that can create ordered magnetic structures even in the presence of star-forming and supernova disruptions.
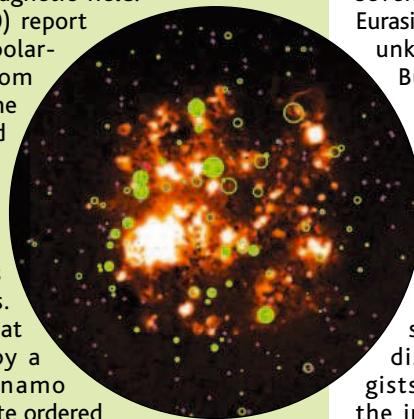
Good Noise

Ambient seismic noise from the atmosphere and ocean collected by seismic arrays is usually discarded by seismologists before they perform the inversion routines that yield crustal structure. **Shapiro et al.** (p. 1615; see the Perspective by **Weaver**) show that cross-correlation of the noise after periods of one or more months can be used to construct higher spatial resolution, three-dimensional images of shear wave speeds. Using data from 60 stations in southern California, the authors produce detailed images of the crustal structure that delineated sedimentary basins from igneous complexes, and even fault lines that offset different rock types. The use of noise has significant advantages for modeling crustal structure and related seismic hazards because it is not necessary to wait for an earthquake to produce seismic waves.

Estrogen Barges In

The steroid hormone estrogen acts both through nuclear receptors that control transcription of target genes, as well as through signaling pathways outside the nucleus. **Revankar et al.** (p. 1625, published online 11 February 2005; see the Perspective by **Hewitt et al.**) report that a G protein–coupled receptor located in the membrane of the endoplasmic reticulum mediates estrogen signaling in various cell types. Upon binding to estrogen, the

receptor stimulates mobilization of intracellular calcium and synthesis of nuclear phosphatidylinositol 3,4,5-trisphosphate, both of which trigger further signaling events. Estrogen is a membrane-permeable molecule, and it is likely that its access to intracellular membrane receptors can facilitate some of the rapid nongenomic signaling initiated by the hormone.

Helping Notch on Its Way

Notch proteins act as receptors for a conserved signaling pathway affecting numerous cell fate decisions, and fucosylation of the glycans on Notch are thought to be important for its function. **Okajima et al.** (p. 1599, published online 3 February 2005; see the Perspective by **Lowe**) find that the fucosyltransferase, OFUT1, in addition to promoting fucosylation of a variety of substrates, including Notch, has a separable Notch-specific chaperone activity. It appears that OFUT1 binds to newly synthesized Notch receptors in the endoplasmic reticulum, where it promotes folding and thereby secretion of the Notch receptor. It is this chaperone function, not the ability to fucosylate the receptor, that is important in maintaining Notch function. It is possible that other glycosyl transferases may play similar roles in the quality control of other membrane and secretory proteins.

Unraveling Signaling Networks

Understanding complex signaling networks is a difficult task that requires new and improved technology. **Barrios-Rodiles et al.** (p. 1621) describe a method of tagging proteins that allows comprehensive mapping of interactions of suspected signaling proteins. High-throughput execution of more than 10,000 experiments yielded a signaling network activated by transforming growth factor β (TGFβ) with more than 900 interactions. The dynamic nature of the network involved connections being both lost and gained as cells respond to TGFβ, which regulates the epithelial to mesenchymal transition that occurs during development and also contributes to invasive properties of carcinomas. **Ozdamar et al.** (p. 1603) explored the regulation of tight junctions by TGFβ and the role of the polarity protein, Par6. Phosphorylation of Par6 by the TGFβ receptor was required for epithelial to mesenchymal transition of mammary gland cells. The function of Par6 appears to be recruitment of an E3 ubiquitin ligase (Smurf1), which leads to degradation of the small guanosine triphosphatase RhoA and dissolution of tight junctions.
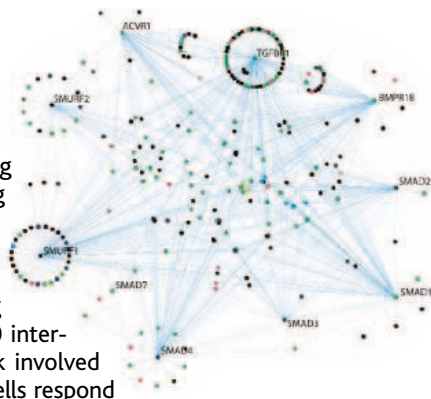
Brassinosteroid Signaling Pathway

Plants lacking a type of steroid—brassinosteroid—are likely to be dwarfed with curled leaves and exhibit an ineffective growth pattern in the dark. Brassinosteroids bind to receptors at the plant cell surface and initiate a signaling cascade that involves nuclear factors including BZR1 and BZR2. **He et al.** (p. 1634, published online 27 January 2005; see the Perspective by **Sablowski and Harberd**) have now characterized aspects of the signaling pathway for brassinosteroids in detail and find that BZR1 is a DNA binding protein that functions as a transcriptional repressor.

Linking Responses to Reward

If the size and probability of rewards are variable, efficient neural coding would argue that our responses would be adjusted to center somewhere in the mid-range of possible reward magnitudes and that the response would be modulated to take into account how wide the range of probable rewards is. **Tobler et al.** (p. 1642) present data that suggest these adjusted responses are in fact encoded within the patterns of activity of dopamine neurons in monkeys as the animals adapted to a schedule of varying rewards.

Confusion at the Space Agency

You would have thought there might be a little joy at the U.S. National Aeronautics and Space Administration (NASA) after the fiscal year 2006 federal budget was released last month. With the success of the Mars rover missions, NASA's space scientists gained the astronomical equivalent of rock star status, and the agency's modest budget increase of 2.4% was four times better than the average for government R&D. But instead, the mood is an odd combination of confusion, gloom, and struggle. What's going on over there?

It starts with two problems. Long before anyone started thinking about the 2006 budget, NASA officials were struggling with what to do about the Hubble Space Telescope. Send astronauts up to fix it? No, said NASA chief Sean O'Keefe, as he left office; too risky. Wrong, said a National Academies panel. A robotic fix is too costly, and a human servicing mission is safe enough. Other proposals were floated, including one for a new telescope that could look for dark energy and dark matter. The president, perhaps feeling saturated by all of this, didn't include servicing money in his budget, leaving scientists to debate priorities.

In fact, priorities and the willingness to set them constitute the second problem. Many astronomers want to see Hubble fixed, or a new telescope put in its place, but they don't want to see money sucked away from other projects. But that's the small end of the NASA problem. On 14 January 2004, President Bush announced a "vision" for space exploration: a project that would take astronauts to the Moon to establish a base and then launch a manned probe to Mars. This announcement, strangely absent from the State of the Union message a week later and still undiscussed in Congress, had a major impact on the NASA budget. According to O'Keefe, it produced a windfall that made the 2006 budget request better than it might have been. But the joy is confined, because the new budget justifies the fears of NASA scientists that exploration will take away funding originally destined for other projects.

At the moment, it appears that with the near-death of the Jupiter Icy Moons Orbiter, there will be no further major robotic explorations of the outer solar system, except the Pluto probe. Considering the scientific haul from the spacecraft Cassini's Saturn sojourn, that's a tragedy. Joining the legion of projects on hold will be the Space Interferometry Mission, which hoped to explore for Earth-sized planets, and the Beyond Einstein project, involving multiple spacecraft arrayed to test the theory of relativity. In short, the imperative of the 3M (man-Moon-Mars) vision has shunted several robotic projects off onto a siding.

The 3M vision may be good news for lunar and martian research, but it is bad news overall for science. Getting humans to Mars is likely to capture public enthusiasm and will require good science and technology. But this is no reason to abandon robotic flights to explore other planets and moons or probe the secrets of deep space. Establishing scientific priorities is difficult enough, given the abundance of technological resources and experimental possibilities available at NASA. Introducing a brand-new exploration mission without additional funding overturns the priority applecart and leaves complex and exciting plans in limbo. That's where NASA is now.

What should be done? First, there's a need for leadership. The president should quickly appoint a new administrator for the space agency, who could unblock the Hubble logjam by following the National Academies' recommendation and ordering a servicing mission. If that doesn't happen, we can expect a continuing argument over alternatives (new Hubble, repaired old Hubble, no Hubble fix at all), with no action. It will help morale and future programs if that decision does not take money from other programs.

Next, the new boss should plead for strong science support from Congress and make it clear that the new exploration program will not be made a reality by raiding existing science money. NASA's science reorganization last summer has left some unfortunate lingering ambiguities. The future of Earth-observing missions is undefined. NASA's hope that the National Oceanic and Atmospheric Administration (NOAA) would take over its research satellites is apparently vain, because NOAA doesn't have the money. The environmental sciences need an effective and successful Earth-observing system, and NASA's new leadership should stand up for that need.

<div align="right">

Donald Kennedy
Editor-in-Chief

</div>
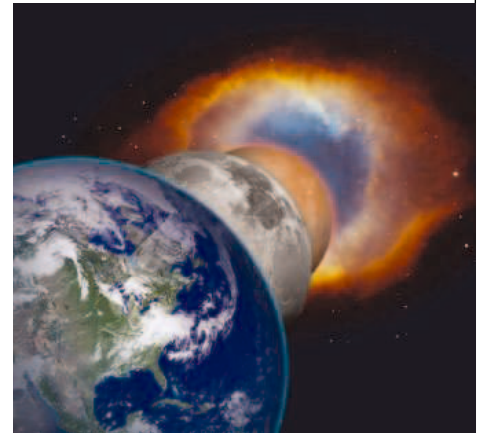
ECOLOGY/EVOLUTION
Sons and Daughters

Biases in the ratio of males to females occur in many polygynous mammal species. According to the mother's condition, investment in sons or daughters may have different fitness benefits in terms of the quality of offspring and hence quantity of grand-offspring produced. In many cases, such as red deer in Scotland, mothers in good condition differentially invest in sons, because males are more costly to rear. However, the reverse may sometimes be true. Kruger *et al.* studied sex-ratio variation over 30 years in a population of springbok in the southern Kalahari region of South Africa. Females in better condition produced more daughters than sons. It seems that the faster onset of sexual maturity in females will produce greater fitness returns in the unpredictable Kalahari environment. Rainfall may be an important controlling factor: Daughters were differentially produced earlier in the wet season, giving them a greater chance of reaching maturity in good condition themselves. The mechanism of sex-ratio adjustment probably lies either in an ability on the mother's part to discriminate between X- and Y-bearing sperm or condition-dependent selective implantation of male or female embryos. — AMS

Proc R. Soc. Lond. B **272**, 375 (2005).

Male springbok.

MATERIALS SCIENCE
Nylon-Nanotube Fibers

One potential application of carbon nanotubes is as a reinforcing agent for polymer fibers, and direct mixing has led to some significant improvements in tensile strength and Young's modulus. However, incomplete dispersion of the nanotubes, which tend to bundle together, and a lack of direct bonding to the polymer, which helps prevent pullout, have limited performance. Gao *et al.* have overcome some of these difficulties by using caprolactam as both solvent and monomer for incorporating single-walled nanotubes (SWNTs) into a nylon-6 matrix. Nitric acid–treated SWNTs, which are terminated with carboxylic acid groups, are well solvated by amide-containing compounds such as caprolactam. After nylon-6 is formed by the ring-opening polymerization of caprolactam, the amino end of the nylon chain can couple to the SWNTs via an amide linkage. The tensile strength and Young's modulus of nylon-6 improved by about a factor of 2 to 3 for SWNT loadings of 0.5 to 1.5 weight %. — PDS

J. Am. Chem. Soc. 10.1021/ja446193 (2005).

MATERIALS SCIENCE
Broken Teeth

Teeth are made up of two calcified tissues that have very different properties: enamel and dentin. The outer coating of enamel is harder but more brittle than the dentin it surrounds. The interface zone between these two structures has been thought to prevent cracks in the enamel from traversing into the dentin, which would cause the fracture and complete failure of a tooth. Using interfacial fracture mechanics, Imbeni *et al.* show that the thin interface layer is not responsible for crack arrest. By creating a series of Vickers microhardness indents in polished sections of healthy extracted teeth, they were able to observe the angle and depth penetration of the cracks that formed. In a majority of the cases, the crack penetrated into the dentin, where it was stopped by the bridging links that form between its mineral and biological components. Although the interface itself is not that strong, the dentin near the interface has collagen fibers that are preferentially oriented perpendicular to the interface and also has a lower mineral content relative to the bulk material, and it is this combination of factors that stops the cracks in their tracks. — MSL

Nature Mat. **4**, 229 (2005).
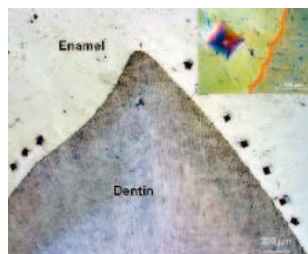
Cracks induced at the enamel-dentin boundary.

VIROLOGY
Virus-Directed Damage Control

Viruses are successful pathogens because of the many and varied ways they usurp host proteins for their own gain.

Uracil DNA glycosylase (UNG2) is part of the base-excision repair (BER) machinery that helps preserve the integrity of cellular DNA. UNG2 is packaged into the virions of human immunodeficiency virus (HIV) type 1, but the enzyme's role in this context is unclear. Priet *et al.* now show that the virion-associated UNG2 is essential to the viral life cycle. UNG2 counteracts the misincorporation of uracil into viral DNA, an event that could be deleterious to the virus. Intriguingly, in experiments exploring the effect of HIV on host BER, Aukrust *et al.* find that CD4$^+$ T cells from HIV-infected patients exhibit a decline in DNA glycosylase activity and are impaired in their capacity to repair cellular DNA damage. Both abnormalities were ameliorated by antiretroviral drugs. Whether or not these effects on BER are mechanistically linked, it's clear that in both scenarios the advantage goes to the virus. — PAK

Mol. Cell **17**, 479 (2005); *Blood* 10.1182/blood-2004-11-4272 (2005)

BIOMEDICINE
Presentable Enough for Entry

In the autoimmune condition multiple sclerosis, demyelination and axonal damage ultimately result in impaired motor function. The disease is thought to be caused by invading T cells that react against self components of the central nervous system (CNS), although the identity and location of antigen-presenting cells (APCs) that activate pathogenic T cells is a matter of speculation.

Greter *et al.* studied a multiple sclerosis system in which T cells reactive to a myelin antigen induce experimental autoimmune encephalomyelitis (EAE) upon transfer to mice. Animals lacking organized central lymphoid tissue developed EAE as quickly and with the same severity as control animals, suggesting that pathogenic T cells do not need to be reactivated in peripheral lymphoid organs in order to migrate to the CNS. Resident APCs of the CNS—microglial cells and astrocytes—did not appear to be important for causing disease. Instead, a subset of nonresident dendritic cells was required for disease to progress. In the model and in multiple sclerosis lesions, similar dendritic cells were associated with microvessels of the CNS, suggesting that activation and entry of autoreactive T cells may occur through the presentation of antigen at the blood-brain barrier. — SJS

Nature Med. **11**, 328 (2005).

GEOPHYSICS
The Sum of the Parts

Quantifying how emissions of any particular greenhouse gas affect the radiative forcing of climate is difficult, because of the complexity of the chemical interactions between different species and the wide range of spatial and temporal scales of atmospheric processes. Current assessments of climate change assume that a particular amount of radiative forcing cannot be attributed to any specific emissions species, and instead rely on calculations based on the atmospheric abundance of each species. Shindell *et al.* use a coupled chemistry-aerosol-climate model to hindcast atmospheric composition from preindustrial times to the present, caused by increased emissions of methane and the precursors of tropospheric ozone (NOx, CO, and volatile organic compounds, excluding methane). The global annual average composition response to all emission changes is nearly the same as that of the sum of the responses to individual emissions. Thus, emission figures can be used to calculate the radiative effects of these species. This emissions-based view indicates that the relative importance of various emissions is significantly different than suggested by current abundance-based assessments: Methane, in particular, is almost twice as important as previously suggested. — HJS

Geophys. Res. Lett. **32**, L04803 (2005).

HIGHLIGHTED IN *SCIENCE*'S SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

Science's stke
www.stke.org

Checkpoint Control at the Golgi

Organelles, such as the Golgi apparatus, must disperse equally during cell division. However, it is not clear whether checkpoints exist for sensing organelle integrity during mitosis. Preisinger *et al.* examined the link between Golgi morphology and cell cycle control. GRASP65, a structural component of Golgi membranes, is required for Golgi fragmentation before entry into mitosis. The C terminus of GRASP65 is phosphorylated primarily by the mitotic kinase Cdk1–cyclin B and to a lesser extent by polo-like kinase 1 (Plk1), an enzyme required for normal mitotic spindle function. Phosphorylation of Golgi-associated GRASP65 on the Cdk1–cyclin B consensus sites correlated with entry into mitosis. Plk1 was detected in a complex with GRASP65 and the Golgi protein GM130 in mitotic cell extracts, but only if GRASP65 was phosphorylated by Cdk1–cyclin B, suggesting that the mitotic kinase creates docking sites on GRASP65 for Plk1. When cells were depleted of Plk1, mitotic fragmentation of the Golgi into clusters was decreased. Overexpression of the GRASP65 C terminus delayed entry into mitosis. However, cells expressing a GRASP65 C terminus harboring a mutant that cannot bind Plk passed through mitosis normally. Passage through mitosis may thus depend largely on the influence of GRASP65-associated Plk1 on the Golgi, where it may help to ensure appropriate Golgi fragmentation and thereby equal partitioning into daughter cells. — LDC

EMBO J. **24**, 753 (2005).
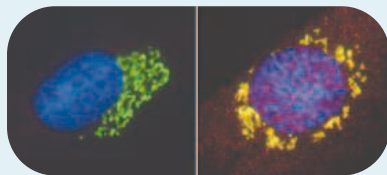
In interphase (left) GRASP (green) labels the Golgi; at the onset of mitosis (right) phosphorylated GRASP (red) also accumulates at the Golgi (yellow) as it starts to disassemble.

CREDIT: FRANCIS BARR, MPI DRESDEN

NETWATCH

edited by Mitch Leslie

COMMUNITY SITE

Racing Light

Particles whipping around inside an accelerator fire off x-rays and infrared and ultraviolet light. Once dismissed as an annoyance, these intense beams now help researchers probe protein structure, gauge the strength of materials, and tackle many other questions. The new site Lightsources.org, sponsored by 17 accelerator facilities, serves as a meeting place for scientists who work with so-called synchrotron light. By paging through a directory, visitors can find out how to sign up for beam time at, say, the Advanced Light Source in Berkeley, California, or the Photon Factory in Japan. The site also rounds up a wealth of resources, from a phone book of European crystallographers to a database for comparing 3D protein structures, and includes a gallery. Above, a diamond-anvil cell for analyzing samples at high pressure.

www.lightsources.org/cms
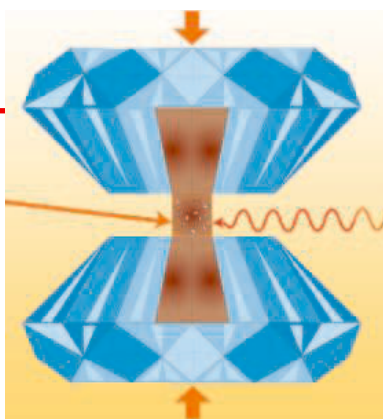
TOOLS

Only Connect

Tracing the interacting molecules that keep a cell running is trickier than keeping track of all the characters in Tolstoy's *War and Peace*. Puzzled readers can turn to Cliffs Notes, while researchers can keep their biochemical networks straight with Cytoscape, a free program for charting and analyzing inter-related genes, proteins, and other molecules. Created by the Institute for Systems Biology, the University of California, San Diego, and other organizations, the software lets users feed in their own data or standard files of molecular interactions from sites such as BIND. The program weaves the information into a map of molecular relationships (right). Cytoscape can also accept data on gene activity determined by microarrays, allowing users to infer hypotheses about which pathway produces a particular gene-expression pattern.

www.cytoscape.org
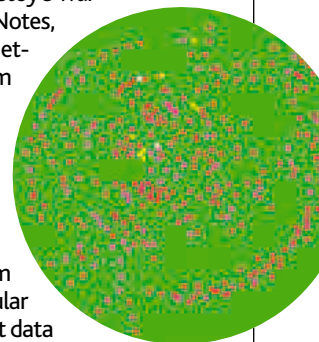
EXHIBITS

The Sum of Human Knowledge

Twenty-six years in the making, the *Encyclopédie* (1751–1777) ranks as one of the intellectual landmarks of the Enlightenment. The work's main editor, the French philosopher and gadfly Denis Diderot (1713–1784), sought not only to summarize human learning but also to foster critical thinking. Thanks to volunteer translators, you can now read more than 100 of the *Encyclopédie* articles at this site from the University of Michigan, Ann Arbor. Translated scientific articles touch on everything from alchemy to probability to the natural history of raccoons. Some entries attempt to reason through questions we're still pondering today, such as whether life exists elsewhere in the solar system. The moon lacks an atmosphere, Jupiter appears too turbulent, and comets undergo temperature extremes, the author concludes: "What living bodies would be able to withstand that extraordinary heat on one hand and extreme cold on the other?"

www.hti.umich.edu/d/did
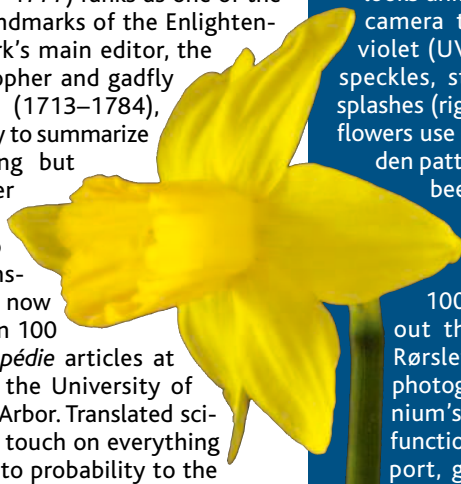
IMAGES

What the Bees See

To our eyes, this narcissus flower looks uniformly yellow (left), but a camera that captures ultraviolet (UV) light reveals speckles, streaks, and splashes (right). Many flowers use these hidden patterns to signal bees and other pollinators, which can detect UV light. For a bee's-eye view of more than 100 plant varieties, check out this gallery from Bjørn Rørslett, a retired water scientist and photographer from Oslo, Norway. A geranium's "bull's-eye" pattern, for example, functions like the runway lights at an airport, guiding approaching insects to a touchdown at the flower's center, where nectar and pollen await.

www.naturfotograf.com/UV_flowers_list.html#top
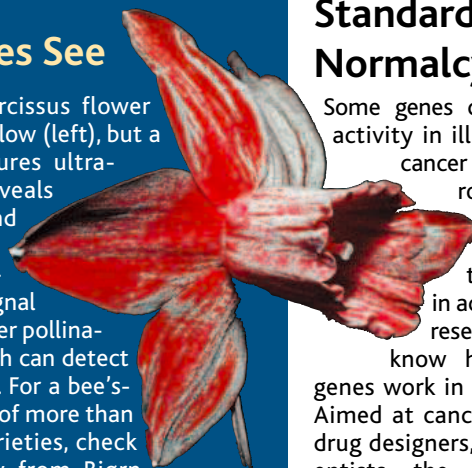
DATABASE

Standard of Normalcy

Some genes crank up their activity in illnesses such as cancer and atherosclerosis, while others get lazy. To identify these changes in activity patterns, researchers need to know how hard the genes work in healthy tissue. Aimed at cancer researchers, drug designers, and other scientists, the new Oncogenomics Normal Tissue Database from the National Cancer Institute provides the baseline data for comparison. After completing a free registration, users can delve into expression results for nearly 19,000 genes in 19 organs, from the adrenal glands to the uterus. The collection caches microarray measurements on fresh tissue samples from apparently hale people who died between the ages of 3 months and 39 years, and the gene roster includes most of the ones that keep cells operating.

ntddb.abcc.ncifcrf.gov/cgi-bin/nltissue.pl

Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

CREDITS (TOP TO BOTTOM): ADVANCED LIGHT SOURCE; CYTOSCAPE; BJØRN RØRSLETT

www.sciencemag.org   SCIENCE   VOL 307   11 MARCH 2005

1539

NEWS

This Week

PAGE 1543
Neutrino bonanza

1545
Early bipedal hominid

BIODEFENSE

Unnoticed Amendment Bans Synthesis of Smallpox Virus

With hardly anyone noticing, Congress has slapped new restrictions—and hefty penalties—on one type of study involving the most dreaded pathogen on Earth. By adding a last-minute amendment to a massive intelligence reform bill in October, Representative Pete Sessions (R–TX) has made it illegal for most U.S. researchers to synthesize the smallpox virus, variola, from scratch. But some virologists, who are only now becoming aware of the amendment, say the law is ambiguous on what exactly is banned, and it could be interpreted to include some research on closely related poxviruses.

By international agreement, only two labs in the world, one in Russia and one in the United States, can store and study variola. U.S. law also criminalizes possession of the virus—along with many other "select agents"—for purposes other than "bona fide" research. But theoretically, nothing has stopped researchers from trying to assemble the virus except for their own conscience.

The new provision, part of the Intelligence Reform and Terrorism Prevention Act that President George W. Bush signed into law on 17 December 2004, had gone unnoticed even by many bioweapons experts. "It's a fascinating development," says smallpox expert Jonathan Tucker of the Monterey Institute's Center for Nonproliferation Studies in Washington, D.C.

Since smallpox was eradicated, the only known variola stocks sit at the Russian State Research Center of Virology and Biotechnology in Koltsovo, Novosibirsk, and the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. But advances in DNA synthesis have made it possible to create viruses in the lab; synthesizing a full, working variola virus

Made to order? It may soon become possible to synthesize variola, the smallpox virus.

may be possible within 5 years, predicts Eckard Wimmer of Stony Brook University in New York, who first synthesized the tiny poliovirus 3 years ago (*Science*, 9 August 2002, p. 1016).

The primary goal of Sessions's amendment —originally introduced as two separate bills, one sponsored by Senator John Cornyn (R–TX)—was to impose much stiffer penalties on the possession of terror weapons, including shoulder-fired missiles, "dirty" bombs, and variola. Until now, for instance, unregistered possession of a select agent carried a maximum penalty of 10 years in prison; under the new law, the minimum is 25 years for variola. Where the law breaks new ground is by also making it illegal to "produce, engineer, [or] synthesize" variola. (Research carried out under the authority of the Secretary of Health and Human Services, who oversees the CDC, is exempt.)

It's extremely rare for the federal government to outlaw specific types of research, ▶

Report Faults Smallpox Vaccination

A review of the ill-fated 2003 U.S. smallpox vaccination campaign charges that the Bush Administration diverged from scientists' advice and moved ahead on a major effort without a clear explanation. The report, issued last week by the Institute of Medicine (IOM), also blames external "constraints" on the Centers for Disease Control and Prevention (CDC) for the program falling short of its goals. CDC Director Julie Gerberding denied the charges.

After the 9/11 attacks and anthrax letters, President George W. Bush in December 2002 announced a plan to vaccinate 500,000 health care workers, and eventually up to 10 million other emergency responders as well as an unspecified number of interested members of the public, against smallpox. But the effort soon foundered, especially after

Ouch. CDC's scientific authority was "constrained" regarding smallpox vaccinations.

the vaccine caused heart problems in a few people, an unexpected side effect. The program wound down in mid-2003, and ultimately only about 40,000 people were vaccinated.

The IOM report[*] notes that "top officials of the executive branch" departed from the recommendations of CDC's vaccination advisory panel, which initially wanted to vaccinate only 20,000 people and later, under political pressure, raised that to 500,000 (*Science*, 20 December 2002, p. 2312). The officials offered "only vague explanation" for vaccinating 10 million more workers and the public, even though the vaccine carried known risks, and there was no evidence of an imminent attack. As a result, workers implementing the program and volunteers expected to line up for vaccinations "remained skeptical," leading to "poor participation," the report says.

The campaign was further hindered because CDC's normally open process of communicating scientific rationale to public health departments "seemed constrained by unknown external influences," the report says. In a strongly worded statement, Gerberding counters that CDC's voice was not "constrained" and that the program "was based on the best scientific advice."

The IOM report refrains from calling the effort a failure. It has apparently improved public health preparedness, as shown by the responses to a subsequent monkeypox outbreak and to severe acute respiratory syndrome, says IOM panel chair and biostatistician Brian Strom of the University of Pennsylvania in Philadelphia. But the panel concluded CDC needs to define and measure smallpox preparedness. Above all, Strom says, while national security concerns have to be balanced against scientific information, CDC "or any other agency needs to speak from the science."

–JOCELYN KAISER

[*] books.nap.edu/catalog/11240.html

Focus

1548
When thoughts are revealed

1551
Lighting up silicon's future
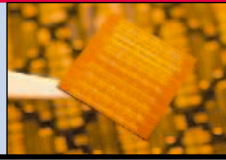
1552
Japan's new drill ship

Pox police. Rep. Pete Sessions introduced stiff penalties for making variola.

says Mark Frankel, who directs the Scientific Freedom, Responsibility and Law Program at AAAS, the publisher of *Science*; the only example he recalls is a 1956 law banning recording or observing jury proceedings, passed in response to certain behavioral studies. To Frankel, the lack of debate about the bill is "worrisome."

Virologists zooming in on the bill's small print, meanwhile, cannot agree on what exactly it outlaws. The text defines variola as "a virus that can cause human smallpox or any derivative of the variola major virus that contains more than 85 percent of the gene sequence" of variola major or minor, the two types of smallpox virus. Many poxviruses, including a vaccine strain called vaccinia, have genomes more than 85% identical to variola major, notes Peter Jahrling, who worked with variola at the U.S. Army Medical Research Institute of Infectious Diseases in Fort Detrick, Maryland; an overzealous interpretation "would put a lot of poxvirologists in jail," he says.

Bernard Moss of the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, believes the word "derivative" means that existing orthopoxviruses are allowed, even if they are highly similar to variola. But on the other hand, the definition does not seem to prevent researchers from taking another poxvirus and adding genes to make it more like variola. "That seems to leave a bit of a hole," Moss says. "It's a funny definition, and it should certainly be clarified," says Paula Traktman of the Medical College of Wisconsin in Milwaukee. A spokesperson for Sessions said that the amendment was "a collaborative effort between the executive and the legislative branches" with "many sources of input" but did not know who had provided the variola definition.

–MARTIN ENSERINK

SPACE SCIENCE
NASA Plans to Turn Off Several Satellites

NASA intends to stop operating more than a half-dozen existing science probes at the end of this year, including the famed Voyager 1 and 2 spacecraft now racing toward the edge of the solar system. Although space agency officials say no final decisions have been made, the agency's 2006 budget request includes no money for a host of solar and space physics projects that currently cost a total of $23 million annually.

In a 2003 speech marking the 100th anniversary of the Wright brothers' flight, President George W. Bush praised the Voyager missions, launched in 1977, as a prime example of "our skill and daring" in exploration. "If the U.S. wants to explore, then turning off Voyager is exactly the wrong signal to send," says William Kurth, a space physicist at the University of Iowa in Iowa City. NASA spokesperson Dolores Beasley says that "Voyager is not canceled," although no funding is planned beyond 1 October.

Voyager 1 is currently 95 astronomical units (AUs) from Earth and may have already passed through the termination shock that marks the solar system's boundary with interstellar space. Physicists are eager to understand what happens when the solar wind ceases and deep space begins, and additional data from Voyager 1 and Voyager 2—which is now 76 AUs from Earth—could resolve the debate over whether Voyager 1 has passed that point. NASA spends $2 million a year to operate the two spacecraft, which are thought capable of transmitting data for another 15 years. "It will be a great loss to shut Voyager off," says Edward Stone, former head of the Jet Propulsion Laboratory in Pasadena, California, which operates the mission.

Voyager is not the only casualty in the 2006 budget plan. NASA also has not budgeted money for five other solar physics missions: the 1997 Transition Region and Coronal Explorer, the 1996 Fast Auroral Snapshot Explorer, the Wind mission launched in 1994 to examine the solar wind, the 1996 Polar to examine the upper atmosphere, and the 1992 Geotail to study Earth's magnetic field. The space agency would also stop funding its portion of the 1990 European Ulysses mission to study the sun. In addition, NASA plans to halt funding for the 4-year-old Thermosphere, Ionosphere, Mesosphere, Energetics, and Dynamics mission at the end of 2006, as well as for the U.S. portion of the European Cluster mission to study the solar wind, which last month was extended through 2009.

Daniel Baker, a solar physicist at the University of Colorado, Boulder, and a member of the National Academies' space studies board, says he is appalled by NASA's decision. He worries that the result will be a lengthy gap in coverage and a dearth of graduate students to seed a new generation of scientists. Margaret Kivelson, a planetary physicist at the University of California, Los Angeles, and also a space studies board member, sees the move as a sign that NASA is willing to sacrifice science projects for Bush's explo-
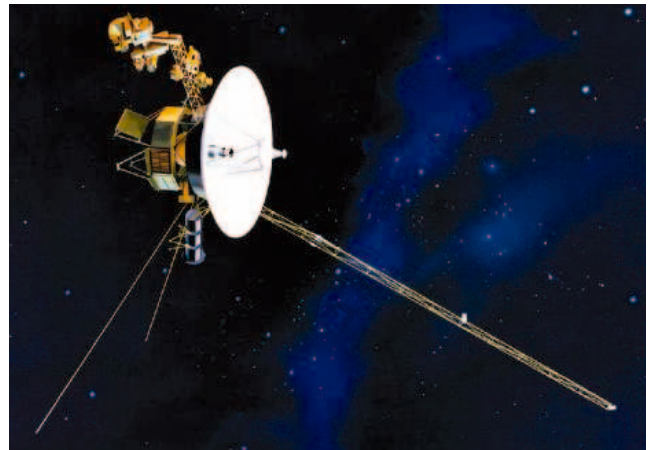
So long, Voyager? NASA may not have money next year to operate Voyager and several other science missions.

ration vision to focus on the moon and Mars.

Beasley says that NASA will review the space missions next month. "Just because the budget says zero [funding] does not mean they will not be getting money," she added. One congressional aide who has begun hearing from worried scientists says that the space agency shouldn't expect to turn off the probes without a fight. **–ANDREW LAWLER**

CREDIT (BOTTOM): NASA

NEUTRINO PHYSICS
Fermilab Experiment Shoots the Muon

BATAVIA, ILLINOIS—Nobody along the 700-kilometer beamline will notice the trillions of particles zooming underfoot—but scientists are certainly taking notice. Last week, a new experiment at the Fermi National Accelerator Laboratory (Fermilab) began sending neutrinos from an accelerator here to a detector deep underground in a Minnesota iron mine. Physicists working on the detector, known as NuMI/MINOS, have high hopes that the experiment will soon eclipse a similar one in Japan and put the most stringent limits on several properties of the mysterious neutrino.

Bull's-eye. Steel plates in an underground lab in Minnesota are designed to capture neutrinos from Fermilab, 700 kilometers away.

"Within a few years of running, we should have of the order of 10,000 events," says Stan Wojcicki, co-spokesperson of MINOS, referring to particle detections. For comparison, the previous best long-distance neutrino-beam experiment, the Japanese K2K, has seen roughly 100 events in the past 6 years (*Science*, 2 November 2001, p. 987). "By summer, we may have a result comparable or even better than K2K," he adds.

At a ceremony at Fermilab last week, Speaker of the House Dennis Hastert (R–IL) officially launched the experiment. "With the launch of this project, Fermilab has positioned itself for the future," he said, shortly before pressing a button on the laptop and getting NuMI/MINOS under way.

NuMI refers to a beam at Fermilab that creates muon neutrinos—nearly massless elementary particles that occasionally change varieties (or "oscillate") into other flavors of neutrino. To create these neutrinos, scientists divert high-energy protons, which ordinarily feed the Tevatron atom smasher, and send them to a graphite target.

The protons hit the graphite, creating pions, which are then focused into a beam by two magnetic horns and release muon neutrinos when they decay. Because neutrinos barely interact with matter, most of the muon neutrinos sail through Earth toward Minnesota and out into space. A few times a day, however, one of them strikes an atom in the MINOS detector—a 6000-ton lump of steel plates with scintillator panels sandwiched in between, shielded from stray particles and cosmic rays by nearly a kilometer of overlying rock. When that happens, the neutrino tends to release a muon, which zooms through a few dozen steel plates before running out of steam. The scintillators flash with light when the muon passes through; by tracking the flashes, scientists can figure out the properties of the neutrino that created it.

Sometimes the beam from Fermilab brings electron neutrinos or tau neutrinos, the results of oscillations. By comparing the number of muon neutrinos produced at the source with the number that reach the Minnesota mineshaft, physicists can figure out how often the muon neutrinos change flavor. This, in turn, reveals the mass difference between two varieties of neutrino, as well as one "mixing angle," a value that describes the fundamental makeup of neutrinos (*Science*, 12 July 2002, p. 184).

Because of the large number of neutrinos produced at Fermilab as well as the bulk and sensitivity of the MINOS detector, physicists believe that NuMI/MINOS will yield orders of magnitude more information about neutrino properties than similar experiments performed in the past. "This is really a new regime in neutrino physics," says Robert Plunkett, deputy project manager for NuMI. "It's a very hot beam. It has to be to do this."

Fermilab's outgoing director, Michael Witherell, says the NuMI/MINOS project, some proposed neutrino follow-ons, and a bid to design and build a huge linear accelerator known as the International Linear Collider (ILC) are the keys to the lab's future. "Neutrinos and the ILC are the headline items," he says. **–CHARLES SEIFE**

Battelle Bows Out of Race to Run Los Alamos Lab

Battelle won't compete for the management contract of Los Alamos National Laboratory, the nonprofit corporation said this week. The withdrawal of the Columbus, Ohio, research giant, which currently manages five Department of Energy (DOE) labs, is good news for the University of California (UC), which has managed the New Mexico facility since its inception 62 years ago. UC is expected to seek another term; other rumored players include Northrop Grumman and General Atomics.

UC's contract expires on 30 September, and DOE plans to release the official request for contract bids shortly. In the wake of complaints from Capitol Hill over the equity of the bid process, DOE recently changed the proposed contract language to require that the new contractor must create a new corporate entity and separate pension fund. The changes, which would dull UC's strengths, have been criticized by New Mexico legislators who want to preserve UC's generous retirement benefits.

Nevertheless, says a former Los Alamos manager, "at this stage UC is still the big entity." **–ELI KINTISCH**

Italian Science Agency Gets Revamp

ROME—A sweeping overhaul of Italy's main science funding agency—the National Research Council (CNR)—will give the system "a more structured approach" and align scientists' work with national goals, research minister Letizia Moratti told *Science* this week. The changes, due to take effect at the end of this month, will group all existing research under 85 "strategic programs." Scientists say they're concerned that the scheme will favor applied research, especially projects endorsed by industry. Moratti insists that fundamental science will be protected, noting that the Berlusconi government has put investigator-driven research on a permanent legal foundation. But some CNR scientists and officials are furious with the new layers of bureaucracy and centralization of power.

Headed by Fabio Pistella, who took office last autumn, CNR will get increased power in its 11 central departments, which will oversee the 108 individual institutes of the old CNR. Contrasting this approach to the U.S. model, one high-level source commented that it "would be unimaginable" for the government to tell the National Science Foundation "what to do."

–SUSAN BIGGIN AND JACOPO PASOTTI

GENE THERAPY

Panel Urges Limits on X-SCID Trials

A U.S. advisory committee last week recommended limits on gene therapy trials in light of a third case of leukemia in a study in France. The panel suggested that U.S. studies of the same disease, X-linked severe combined immunodeficiency (X-SCID), should enroll only patients for whom conventional treatment has failed. However, trials of related diseases, as well as gene therapy trials using similar retroviral vectors, should continue, the panel said. The third leukemia "doesn't change the sense of unease dramatically," said chair Mahendra Rao of the National Institutes of Health (NIH).

Gene therapy trials for SCID have been the field's only success; since 1999 gene therapy has restored the immune systems of at least 17 children with two forms of the disorder. Excitement turned to worry in late 2002, however, when two children developed T-cell leukemia in a trial of X-SCID led by Alain Fischer at the Necker Hospital in Paris; one

child died last fall. Although trials put on hold later resumed, a report that a third child in the French trial developed leukemia in January rekindled concerns about the therapy's risks (*Science*, 18 February, p. 1028).

This latest leukemia appears to be different from the previous two. Those occurred after a retrovirus carrying a gene called *gamma c* inserted into the oncogene *LMO2* in bone marrow cells in infants less than 3 months old, noted

Success story. Christopher Reid, a patient in a British X-SCID gene therapy trial.

Food and Drug Administration (FDA) official Carolyn Wilson at a meeting of the FDA Cellular, Tissue, and Gene Therapies Advisory Committee. According to data provided by

Fischer and French authorities, the third child, who was treated at 9 months old, does not appear to have an *LMO2* insertion. Although the vector again apparently landed on an oncogene or oncogenes, the insertions occurred at three sites that have not yet been identified.

The panel also heard other new data, which offered a mixed message. Last September, a monkey died from a leukemialike cancer at NIH, apparently as a result of being treated in 1999 with a retrovirus carrying two marker genes, reported Cynthia Dunbar of NIH. On the other hand, NIH's Utpal Davé described a report last year in *Science* on a retrovirus-induced mouse ▶

INDIA

Prime Minister Backs NSF-like Funding Body

NEW DELHI—Indian Prime Minister Manmohan Singh has endorsed the creation of an independent agency to support basic research—with a proposed budget that's more than three times the amount the government is now spending.

Scientists have long complained about the current process for winning grants, including inflexible rules and funding decisions that take more than a year. Last week Singh attended the first meeting of the new Science Advisory Council to the Prime Minister and embraced its recommendation for a National Science and Engineering Research Foundation with a mandate to "strongly promote and fund research in all fields of science and engineering." The new foundation "is being patterned on the lines of the acclaimed U.S. National Science Foundation," says C. N. R. Rao, chair of the council, who has campaigned for more than a decade for such a freestanding body. "A foundation that manages its own accounts and is run by a scientist is the only hope for reversing the rapid decline in Indian science," he adds.

The council recommended an annual budget of $250 million for the foundation. That

amount would dwarf the $72 million now being spent by the Science and Engineering Research Council (SERC), an arm of the Department of Science and Technology (DST). The management and operating structure of the new foundation would be familiar to most U.S. scientists: five research directorates and a part-time body of distinguished scientists setting its overall direction. The council also recommended that the new foundation be responsible for "assessing the overall health of Indian science" (as NSF does with its biennial *Indicators* report) as well as

funding "units of excellence [run by] researchers of exceptional merit" (as NSF does with centers focused on particular research areas).

An evaluation of the existing structures by the prime minister's council was sharply critical of SERC, which was founded in 1972 and supports the bulk of fundamental research done in India. "Science funding in academic institutions and universities has not kept pace with the growing costs of basic research," it concludes. Instead, the process has become "mired in bureaucracy, with complex financial procedures inhibiting efficient operation." Even so, the secretary of DST, Valangiman Subramanian Ramamurthy, say he "has no objections to the new body, since the basic idea is not bad."

Science Minister Kapil Sibal has been asked to work out the details, including the fate of SERC. "There is no question of anybody saying no when the prime minister has said 'Yes, it must be set up,'" says Sibal. The change can't come too soon for Rajendra Kumar Pachauri, director general of The Energy and Resources Institute in New Delhi. "An independent foundation," he says, "is vital for resuscitating … a moth-eaten" scientific establishment.   **—PALLAVA BAGLA**

A solid foundation. Prime Minister Singh is flanked by top science aides Kapil Sibal (left) and C. N. R. Rao (right).

leukemia that contained insertions in both *LMO2* and *gamma c*, the gene corrected by the X-SCID therapy (*Science*, 16 January 2004, p. 333). The two genes seem to "cooperate" in causing cancer, Davé said, suggesting that gene therapy for diseases not involving *gamma c*—which itself may be oncogenic when expressed by a retrovirus—may be safer.

Indeed, panelists noted, no leukemia cases have yet been seen in trials of ADA-SCID, which does not involve the *gamma c* gene. Nor have leukemias appeared in an X-SCID trial in the United Kingdom that has treated 7 patients. However, the French leukemias appeared roughly 33 months after treatment, and the U.K. patients have not reached that point.

The panel concluded that if two X-SCID trials now on hold in the United States resume, they should enroll only children who have failed bone marrow transplants. "That's going to be a very small number," said panelist Daniel Salomon of the Scripps Research Institute in La Jolla, California. But the panel suggested FDA could lift its hold on a U.S. trial for ADA-SCID. Researchers will be watching closely to see whether any leukemia cases turn up in the British trial. If not, "that would certainly change things" because it would suggest conditions specific to the French trial are leading to the leukemias, concluded Rao.

–JOCELYN KAISER

PALEOANTHROPOLOGY
Skeleton of Upright Human Ancestor Discovered in Ethiopia

Scientists working in the remote badlands of Ethiopia have found the oldest known skeleton of an upright walking hominid, roughly dated to nearly 4 million years ago. The remarkably preserved partial skeleton includes many bones of the pelvis, leg, back, and arms, as a team led by paleoanthropologists Yohannes Haile-Selassie and Bruce Latimer of the Cleveland Museum of Natural History in Ohio announced last week at a press conference in Addis Ababa, Ethiopia.

The shape of the top of the lower leg bone and pelvis have already convinced the discoverers that this hominid walked on two legs, which is the traditional hallmark of being a member of the human family rather than an ancestor of apes. "It's a once-in-a-lifetime discovery," says Haile-Selassie.

The skeleton so far also includes precisely the anatomical parts below the neck that can allow scientists to distinguish whether it

Early walker. The owner of this shinbone walked upright in Ethiopia 4 million years ago.

CREDIT: ANTHONY MITCHELL/AP PHOTO

walked like a modern human or in a more primitive manner. "It's a monumentally important skeleton, a real key to understanding hominid origins," says paleoanthropologist Carol Ward of the University of Missouri, Columbia, who cautions that she has not seen the as-yet-unpublished skeleton. "The bits from the skeleton are exactly the pieces we need to see if we came from something like a chimp or something more primitive."

The skeleton was found on 10 February near the village of Mille in the central Afar Depression, where a sharp-eyed fossil hunter named Alemayehu Asfaw spotted an elbow bone. Soon team members found the other part of the arm bone, the pelvis, leg bones, ribs, vertebrae, clavicle, and scapula. Extinct pigs found with the skeleton suggest that it lived 3.8 million to 4 million years ago, a critical time when humans were evolving the ability to walk. The team is now dating samples of volcanic rock taken from layers above and below the fossil and studying fragmentary fossils, including leg and toe bones, from 11 other individuals.

The identity of the new skeleton is still unclear, in part because the specimens are still embedded in matrix and also because most of the known fossils of this age are so fragmentary. There are only four other partial skeletons of human ancestors older than 1 million years. Contenders for the new skeleton's identity include the slightly younger *Australopithecus afarensis*, whose most famous member is Lucy, a partial skeleton that lived 3.2 million years ▶

ago at Hadar, 60 kilometers south of Mille. An older Kenyan species thought to be bipedal, 4.1-million-year-old *A. anamensis*, is also a possibility. Haile-Selassie says the new skeleton is slightly younger and distinct from the mysterious 4.4-million-year-old *Ardipithecus ramidus*, known from teeth and a crushed, still unpublished, skeleton that he also found; he adds that the new skeleton may connect the dots between

Ardipithecus and later australopithecines, revealing how the human mode of walking evolved. Three even earlier species have been proposed as bipedal hominids but are known only from fragmentary fossils or a skull.

The discovery of the new skeleton comes at a good time for Haile-Selassie, one of the first black Africans to launch his own fossil-hunting expedition (*Science*, 29 August 2003, p. 1178).

The U.S. National Science Foundation rejected his grant application last year to look for hominids in the localities around Mille. Instead, he and Latimer got foundation funding for a small team of mainly Ethiopian fossil hunters. With a find like this, Haile-Selassie hopes getting future grants will not be a problem. "We want to go out and see if we can find the head and mandible," he says.   **–ANN GIBBONS**

ALZHEIMER'S DISEASE

Play and Exercise Protect Mouse Brain From Amyloid Buildup

As the population ages, finding ways to stave off the debilitating brain degeneration of Alzheimer's disease becomes ever more critical. New results with a mouse model of the condition now provide further support for the idea that "use it or lose it" applies as much to the mind as to the body.

A leading explanation for Alzheimer's disease blames abnormal buildup of a small protein called β amyloid, which accumulates in pathological structures called plaques in patients' brains. Now, working with mice genetically engineered to produce similar β-amyloid plaques, a research team led by Sam Sisodia of the University of Chicago, Illinois, has found that the β-amyloid buildup can be greatly reduced by a lifestyle change: housing the animals in an enriched environment—one amply stocked with toys and exercise equipment—instead of in standard lab cages equipped with nothing more than food, water, and bedding material.

The experiments, reported in today's issue of *Cell*, also provide clues to how an enriched environment might protect against β-amyloid accumulation. Zaven Khachaturian, editor of the journal *Alzheimer's and Dementia*, calls the work "very provocative. … It opens new ways of getting at the underlying mechanism" of plaque formation.

Several epidemiological studies have suggested that environmental enrichment, including education and intellectually challenging leisure activities such as reading and playing bridge, diminishes the risk of Alzheimer's disease. Others have pointed to a possible protective role of exercise. But lower activity levels could be an early symptom of the disease rather than a risk factor.

With mice, though, it's possible to study environmental influences on the earliest stages of plaque formation. Sisodia and his colleagues Orly Lazarow and John Robinson started their experiments when the mice were just 1 month old, many weeks before they nor-

mally show symptoms of Alzheimer's disease; the genetically modified animals they used ordinarily develop β-amyloid plaques by about 4.5 months of age. The researchers put seven animals in standard cages and another nine in the enriched environment, where the activities of the mice were closely monitored.

After 5 months, the researchers killed both sets of mice and examined their brains. Animals kept in the enriched environment showed "a marked reduction in amyloid burden," Sisodia says. The decrease appeared to be related to exercise. "The animals that were most active as

Fun and games. Mice in cages with toys and exercise equipment develop less β amyloid than do ones in standard cages (*inset*).

determined by their time on the running wheels had the least [β-amyloid] burden," Sisodia adds. He notes, however, that other aspects of the enrichment, such as increased visual stimuli and social interactions, could still account for the reductions.

The researchers also identified changes in the brain that might explain a lessening of β-amyloid deposition. They saw increased activity of a β-amyloid–degrading enzyme called neprilysin in the brains of the enriched mice, as well as changes in gene expression that could promote neuronal survival and enhance learning and memory.

In late 2003, Joanna Jankowsky of the California Institute of Technology in Pasadena, David Borchelt of the Johns Hopkins University School of Medicine in Baltimore, Mary-

land, and their colleagues reported that enriched environments actually increase plaque formation. The reason for the discrepancy is unclear, although the design of the 2003 experiment was different. For one, that study involved only female mice, whereas the Sisodia team used males. The Jankowsky-Borchelt group also had many more animals in their enriched cages and added young mice as they removed older ones. "To me that spells stress," says David Arendash of the University of South Florida in Tampa, who also studies the effects of enrichment on Alzheimer's mice. That stress might have overcome any beneficial effects of the enhanced environments.

Sisodia's group didn't test whether the enriched cages improved learning and memory in their animals, although work by others suggests that it may. This was the case in the experiments performed by Arendash. The improvement occurred even though the Tampa team did not see reductions in β-amyloid deposition in their mice. But those animals were very old—16 months at the start of enrichment—and they already had extensive β-amyloid deposition.

How much these mouse studies of enriched environments relate to Alzheimer's disease in people remains to be seen. Adding another clue, Constantine Lyketsos and his colleagues at the Johns Hopkins Medical Institutions in Baltimore will report in the April issue of the *American Journal of Epidemiology* that engaging in a variety of different physical activities can reduce the risk of developing Alzheimer's disease by as much as 50%, although only in people who did not carry a particular gene variant called *APOE4* that increases Alzheimer's risk.

Lyketsos says that his team's results and Sisodia's provide an "interesting convergence" about the possible effects of physical exercise on Alzheimer's risk. So while you're out running to save your heart, you might also be saving your brain.   **–JEAN MARX**

Advances in neuroimaging may provide the ability to "read" someone's mind, rightly or wrongly

Brain Scans Raise Privacy Concerns

If you could find out whether those occasional moments of forgetfulness herald an old age ravaged by Alzheimer's disease, would you want to know? Would you want other people to know?

What if tests were available that could determine whether a child could benefit from accelerated classes, whether someone on the witness stand were lying, or if a violent criminal were likely to attack again? Should such tests be used?

None of these tests is available today, and some may never be. But rapid progress in imaging the structure and function of the human brain is forcing neuroscientists and bioethicists to consider the possible consequences of ongoing brain research. The President's Council on Bioethics has launched a series of discussions on neuroimaging and other issues raised by the neurosciences, and the newly dubbed field of neuroethics has received a boost because of concerns about what brain scans might eventually reveal. Many speculations remain uncertain because the ethical quandaries posed by new means of imaging the brain will depend on what those technologies eventually can do. But researchers are already talking about a future in which issues of privacy—keeping information to oneself—and confidentiality—preventing the unauthorized release of sensitive information—loom large.

Triumphs and challenges

Neuroimaging technologies such as positron emission tomography, functional

Laid bare. Neuroimaging techniques may offer a glimpse into the tumult and pandemonium inside someone's head, as this 16th century print by Mattias Greuter suggests.

magnetic resonance imaging, and near infrared spectroscopy have produced wonders in medical clinics and research labs. Physicians have been able to pinpoint damage caused by injuries or illness, and brain scientists have begun to piece together the neural mechanisms involved in perception, cognition, behavior, and emotion.

But the ability to watch the brain in action raises many questions about when, if ever, society has a right to know what someone is thinking. "If some of these technologies become available, it could change how we live enormously," says Henry Greely, a law professor at Stanford University in California who has written extensively about the legal and social implications of neuroimaging technologies. "To the extent that small, easy-to-use devices could tell, either voluntarily or surreptitiously, what was going on inside someone's head, that could have enormous uses throughout society—and also what we today would consider abuses."

Many ethical issues arise from straightforward extensions of current studies. For example, neuroscientist Turhan Canli and his colleagues at Stony Brook University in New York have been examining the correlations between brain scans and personality. Several years ago they showed that when people

Mental State and Characteristics Studied With Neuroimaging: A Sample

consumer preferences

intelligence

racial antipathy

impulsivity

addictive cravings

classified as extroverts on personality tests viewed smiling faces, they tended to have greater activation of the amygdala, a brain region involved in processing emotions (*Science*, 21 June 2002, p. 2191), than did less extroverted people. Since then, Canli and his co-workers have drawn similar connections between personality traits and other subcortical and cortical regions. Meanwhile, other researchers have been linking patterns of brain activity to characteristics such as neuroticism, risk aversion, pessimism, persistence, and empathy.

The links between brain activation patterns and personality are still too tentative to find applications outside the research lab, Canli says. But he points to a number of people who might like to supplement existing sources of information with brain scans, such as school admissions officers, potential employers, or law enforcement personnel. Another worrying possibility, he says, is that a personality assessment could be performed while ostensibly conducting a scan for other reasons, because the person in the scanner could be asked simply to look at pictures or respond to questions.

Beyond personality assessment lies the prospect of detecting defects in brain functioning that could contribute to criminal acts. Imaging studies have shown that moral reasoning engages parts of the brain that are not involved in other forms of reasoning, and other studies have found reduced activity in some of the same brain regions among convicted murderers. One goal of "forensic neuroimaging," says Canli, is to determine whether individuals with a reduced ability to feel empathy, guilt, or remorse about criminal acts exhibit a unique neural signal. If so, this information could be used to monitor individuals at risk of carrying out a criminal act or in sentencing and parole decisions.

A window on thought

Privacy issues are an even greater concern with neuroimaging techniques that can detect ongoing thought processes. In one of the most widely reported neuroimaging studies of recent years, Elizabeth Phelps of New York University, Mahzarin Banaji of Harvard University, and their colleagues used behavioral tests to measure the atti-

Proceed with caution. Law professor Henry Greely says the brain-imaging technologies on the horizon have the potential for enormous good—and abuse.

tudes of a group of European-American research subjects toward African Americans. They then scanned the brains of those subjects while they were viewing unfamiliar African-American faces. Subjects with more negative views of African Americans tended to have greater activation of the amygdala. "I don't think we've gotten to the point where we can say anything about how people will act in the future, but I think we will—it's a matter of time," Phelps says. Other investigators have been looking for distinctive brain activation patterns associated with sexual preferences, political affiliations, and feelings of religious transcendence.

Among the most controversial neuroimaging studies have been those focused on deception. Several research groups have claimed that they can detect brain activation patterns indicative of lying, and one commercial company has begun offering a brain test for deception. Whether these techniques are more reliable than existing

approaches such as polygraphs has yet to be determined. Still, the Defense Department and CIA are sufficiently interested that they have been investing millions of dollars in neuroimaging technologies that might be used in law enforcement or intelligence. A particular focus of this work: brain scans that might reveal the identities of terrorists.

The ability to detect deception reliably could have profound consequences for the legal system, Greely points out. The truthfulness or biases of defendants, witnesses, judges, and juries could be assessed. Entirely new legal procedures might be necessary. For example, if people swore that their testimony was truthful, would the state have the right to test those oaths with brain scans?

The need for perspective

Such scenarios can be chilling, but they also should be viewed with caution, say researchers and ethicists. No one can be sure if any of these possibilities will be real-

political affiliations   aggression   moral reasoning

sexual attraction   empathy

neuroticism   deception

religious transcendence
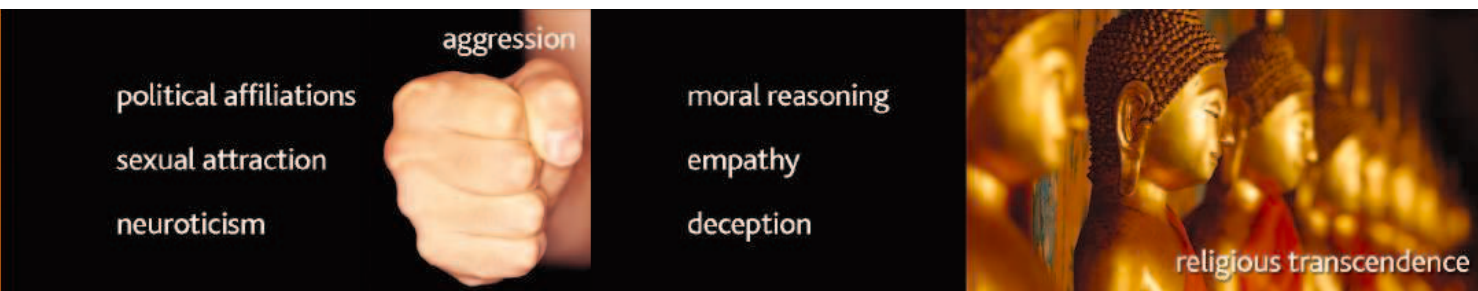
An Image of Disease

Diagnosing diseases through neuroimaging raises issues posed by other biomedical technologies, but often in startlingly personal ways. Consider what neuroscientists call incidental findings. When subjects receive brain scans as part of a research project, the resulting images sometimes bear unwelcome news. According to Judy Illes, who directs the program in neuroethics at the Stanford Center for Biomedical Ethics, 2% to 8% of research subjects turn out to have tumors, malformations, or other clinically significant neurologic problems that were previously undetected.

At a meeting at the National Institutes of Health in January, participating clinicians, researchers, and bioethicists agreed that the possibility of incidental findings should be considered when designing a study and obtaining consent from subjects. Another point of agreement: To protect privacy, the research subject or a surrogate should be the first to hear about a problem, not a physician. But participants could not settle on a standard procedure to detect and respond to incidental findings. Some researchers have every brain scan examined by a radiologist for signs of trouble, whereas others refer only those with obvious abnormalities. "There were areas where the different disciplines had different viewpoints, and those were extremely valuable in understanding the problem and identifying appropriate pathways to solving it," Illes says.

Similar issues arise when a brain scan, advertently or inadvertently, reveals a medical condition for which there is no known treatment. For instance, neuroimaging technologies have proven fairly successful in identifying mild to moderate cases of Alzheimer's disease. But in the absence of a cure, a positive diagnosis may be more of a curse than a blessing. "Are there some things we would be better off not knowing about ourselves? Absolutely," says Martha Farah, the director of the Center for Cognitive Neuroscience at the University of Pennsylvania in Philadelphia.

Getting a brain scan for early signs of Alzheimer's disease is comparable to being tested for Huntington's disease, an incurable neurologic disorder caused by a defective gene. But most genes are several layers removed from our physical or behavioral traits, bioethicists point out. Brain scans, in contrast, tap into mental processes that relate directly to our personalities, our behaviors, and even our private thoughts.

–S.O.

Incidental findings. Research scans sometimes turn up unexpected brain abnormalities, such as this malformation in the right frontal cortex.

ized. For one, current neuroimaging technologies remain expensive and ungainly. "You have to stick someone in a scanner, and they have to be compliant," says Randy Buckner, a neuroscientist at Washington University in St. Louis, Missouri, who helped develop functional magnetic resonance imaging. "It's presently not useful for rapid screening." The equation might change if imaging technologies were no bigger than a set of headphones, or if sensing could be done from a distance, but today such devices remain in the realm of science fiction.

Many questions also surround the validity of brain scans. Some skeptics already refer to neuroimaging as high-tech phrenology, pointing toward poorly designed and impressionistic studies. Others wonder if brain scans can ever match even the accuracy of polygraphs, which use physiological measures of nervousness to detect deception. Polygraph evidence has generally been rejected by all federal courts and state courts except those of New Mexico because of concerns about accuracy. Before neuroimaging could offer useful guidance in the legal system or elsewhere, it would need to be thoroughly tested to see how often brain scans are misleading or incorrect and whether people can train their minds to fool the machines.

Another fundamental question is whether brain scans necessarily reveal information that is not available in other ways. If brain scans are used to draw correlations between neural activation patterns and personality or behavioral tests, why not just rely on the behavioral tests? "We have other ways of finding out how people think about things," says Phelps. "Brain imaging brings another measure of that."

Neuroimagers agree that any brain scan must be compared to an average level of activity, either for an individual or a group. But brain activation patterns differ from person to person and from one instance to another, so measuring departures from an
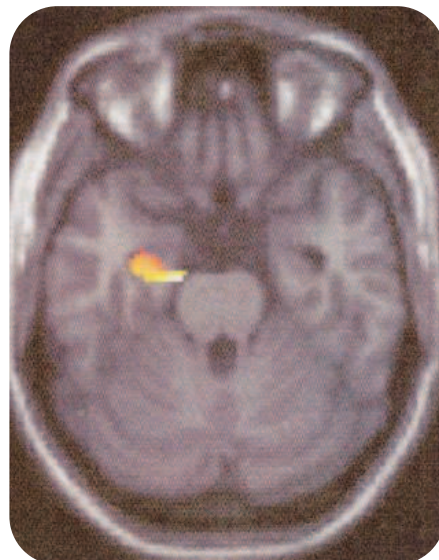
False accuracy? The striking colors and contrast of a brain scan can convey a sense of "objectivity" that may not be warranted, experts caution.

average inevitably involves considerable judgment. Brain scans represent "statistical inferences rather than absolute truths," in Canli's words.

Indeed, researchers and bioethicists alike say that the greatest threat to individual liberty may come not from the capacity of scanners to reveal hidden thoughts but from the mistaken belief that the results of brain scans are highly accurate. The striking colors and contrasts of a brain scan can seem objective or "scientific," even when the appearance of the scan is the product of a technician's image processing. "Probably the only thing worse than having people successfully reading your mind with brain imaging is having people unsuccessfully reading your mind with brain imaging and thinking that they can trust that information," says Martha Farah, who directs the Center for Cognitive Neuroscience at the University of Pennsylvania in Philadelphia.

Maintaining a sense of perspective is important, say researchers and bioethicists. Despite the remarkable technological advances of recent years, human beings are unlikely to give up their secrets easily. As Canli says, "If we could predict what someone will do with 100% accuracy, it would mean that free will doesn't exist—and I'm not prepared to accept that."

–STEVE OLSON

Steve Olson's latest book is *Count Down: Six Kids Vie for Glory at the World's Toughest Math Competition*.

New Generation of Minute Lasers Steps Into the Light

Long-awaited long-wavelength Raman lasers built on microchips are primed to take the next strides in merging light beams and electronics

Microchip-based diode lasers have had a good run. They're at the heart of CD and DVD players, computer disc optical drives, and a host of medical devices. Together, these and other applications add up to a sweet $3.5 billion market. But diode lasers can't do it all. Researchers have struggled to get them to produce the long-wavelength light—ranging from the mid-infrared to terahertz frequencies—that is highly sought after for applications from explosives detection to biomedical imaging. Researchers have also had a tough time making the lasers out of silicon, the workhorse of computer technology, an advance that could vastly improve computer processing speeds by enabling chips within computers and local networks to send signals through high-speed glass fibers instead of metal wires. Now a spate of advances could finally help chip-based lasers leap those hurdles.

In recent months groups at the University of California, Los Angeles (UCLA), and Intel Corp. have reported major strides in making "Raman" lasers out of silicon. Like other lasers, the new silicon-based devices trap light waves, force their peaks and troughs into orderly alignment, and then release them in energetic beams. The one downside is that in order to work, these lasers must be primed by light from another laser. But 2 weeks ago, a group at Harvard University in Cambridge, Massachusetts, reported creating a chip-based Raman laser that works when fed electricity. "Over the past 5 months, this field has exploded," says Philippe Fauchet, an optics expert at the University of Rochester in New York.

The lasers take their name from the Indian physicist Chandrasekhara Venkata Raman, who discovered the principle behind them in 1928. When monochromatic light passes through a transparent material, he found, most of the photons emerge with their wavelength unchanged. Others, though, collide with atoms in the material and lose or gain energy, causing them to emerge at a shorter or longer wavelength.

The effect lies at the heart of fiber optic–based commercial devices called Raman amplifiers, which boost longer-wavelength optical signals streaming through glass fibers for long-distance data transmission and telecommunications. The devices work by using an initial high-energy "pump" pulse to prime the fiber so that when photons in a data pulse pass through, they stimulate the release of additional photons at the same energy, amplifying the pulse. By reflecting the growing light pulse back and forth through a transparent fiber, engineers can cre-
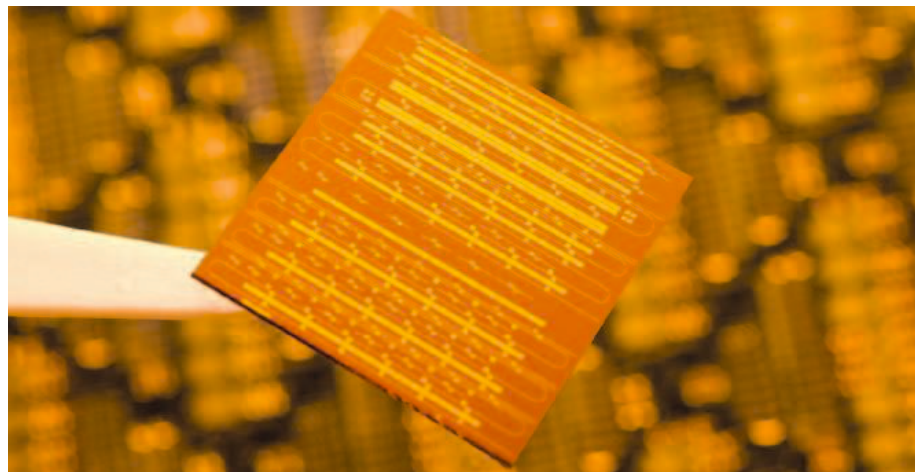
Chip shot. The first continuous-wave silicon laser.

ate a Raman-based fiber-optic laser. But because the Raman effect is so slight in glass fibers, these devices typically require kilometers of fiber to work.

The good news is that the Raman effect is 10,000 times stronger in pure silicon than in glass. "We can do in centimeter-sized devices in silicon what is done in kilometers in glass," says Mario Paniccia, who directs Intel's photonics technology laboratory in Santa Clara, California. At least, that's the theory. Unfortunately, silicon has an appetite for eating laser photons. When an incoming laser pulse—known as the pump pulse—is trained on silicon, silicon atoms can absorb two photons simultaneously. The energy excites one of the atom's electrons, freeing it to roam through the crystal. Such mobile electrons are strong photon absorbers and quickly quench any amplification of laser photons in the material.

Last fall, UCLA optoelectronics researchers Ozdal Boyraz and Bahram Jalali were the first to overcome this problem and create a silicon-based Raman laser. In the 18 October 2004 issue of *Optics Express*, the pair reported that to prevent the buildup of excited electrons, they zapped their silicon chip with a staccato of pulses, each lasting just 30 trillionths of a second, or picoseconds. Between pulses they gave the excited electrons time to relax back to their ground state, so they wouldn't reach a level that kept photons from building up in the material. The UCLA device, however, wasn't pure silicon: It also used 8 meters of optical fiber to carry the emerging laser light back to the silicon crystal for additional passes in order to boost the output of the Raman-shifted pulse.

Three months later, researchers at Intel did away with the optical fiber. In the 20 January issue of *Nature*, a team led by Paniccia reported creating the first all-silicon-based Raman laser. Like the UCLA device, it relied on pulsing an incoming beam, but mirrors in the silicon bounced the light back and forth without the need for the fiber. The Intel team also added another trick: They routed the light down a path within the chip lined with positive and negative electrodes. When the researchers applied a voltage, charged particles swarmed to the electrodes, sweeping the mobile electrons out of the path of the incoming photons. As a result, the team could blast the silicon chip with a stronger pump pulse to increase the output of the Raman-shifted laser light. Last month in *Nature*, the Intel team reported another improvement, the first silicon Raman laser that emits a continuous beam of photons. Boyraz and Jalali jumped back into the fray as well, reporting in the 7 February issue of *Optics Express* that they had incorporated an electric modulator into their optically pumped device to switch their new lasers on and off.

The string of advances, Fauchet says, sets the stage for a host of innovations, such as silicon-based optoelectronic devices to replace copper wires in speeding short-distance communication between computers, as well as other military, medical, and chemi-

cal detection applications. By leveraging the semiconductor industry's decades of experience in fabricating silicon components, the new work could help slash costs for optical components. "It's a potential sea change that allows you to do new things because they are cheap," Fauchet says.

The new lasers have their drawbacks. "The major limitation of Raman lasers is that to get a laser you need another [pump] laser," Fauchet says. "Ideally, you would like to have an electrically pumped laser. That would be the Holy Grail."

As if on cue, in the 24 February issue of *Nature*, researchers led by Federico Capasso of Harvard reported just such a device. Unlike the previous lasers, however, the new one is made from alloys of aluminum, gallium, indium, and arsenic rather than

silicon and works in a different manner. Known as a "quantum cascade" (QC) laser, it consists of hundreds of precisely grown semiconductor layers. As electrons pass through the layers, they lose energy at each step, giving up photons, which combine to create the laser beam.

Capasso and his colleagues at Harvard and Lucent Technology's Bell Laboratories in Murray Hill, New Jersey, had spent a decade building QC lasers that emit light in the mid-infrared range. In hopes of extending their reach to longer, terahertz frequencies, Capasso teamed up with theorist Alexey Belyanin of Texas A&M University in College Station, who had suggested modifying the device by adding new sections that use the Raman effect to shift the initial laser light to a longer wavelength. In essence, the group

created a pair of Raman lasers on a single chip: one that converts electricity into an initial pump laser, and a second that shifts the light to longer wavelengths. The new QC Raman lasers turn out beams of infrared light with a wavelength of 9 micrometers. Capasso says his team is working to create similar devices that turn out beams at terahertz frequencies, which are widely sought after for use in detecting explosives and other chemicals. Fauchet notes that the advance doesn't produce the shorter wavelength photons ideal for telecommunications, but "it demonstrates you don't need an external laser to get a Raman laser," he says.

No matter which of the new Raman lasers proves most successful, the devices look likely to extend diode lasers' run for a long time to come.   **–ROBERT F. SERVICE**

Ocean Drilling

Japan's New Ship Sets Standard As Modern, Floating Laboratory

Scientists expect the *Chikyu*'s massive size and unique capabilities to unlock important secrets that lie underneath the ocean's floor

NAGASAKI, JAPAN—When was the last time scientists got almost everything they wanted? Japan's new riser drilling ship may eventually turn out to have some flaws and limitations. But as the $550 million *Chikyu* nears completion, researchers involved in the 18-nation Integrated Ocean Drilling Program (IODP) can hardly contain their excitement. "They've pretty much done it all," says Richard Murray, a marine geochemist at Boston University and chair of an IODP panel that put together a wish list of instruments for the vessel.

Last month reporters were invited to tour the *Chikyu* as it sat in the Mitsubishi Heavy Industries shipyard here, preparing for a series of shakedown cruises beginning this fall. On display is a vessel designed to drill more than twice as deep as previous drill ships, up to 7 kilometers below the sea floor. It can also work in areas with gas or oil deposits that have been off limits for environmental reasons.

Those capabilities promise a better understanding of key questions such as seismicity beneath the seas, the recycling of oceanic mantle, geologic changes in sea levels, and Earth's climate history. At 210 meters and 57,500 metric tons, the *Chikyu* is 45% longer and 2.4 times the weight of IODP's current workhorse, the *JOIDES Resolution*, and it has 60% more laboratory space, spread over four decks. The labs are now being filled with $18 million worth of equipment, some of

which has never been installed on a drill ship before. "[*Chikyu*] will probably be as well-equipped as the best land-based laboratories

in the world," marvels Mike Coffin, a geophysicist at the University of Tokyo's Ocean Research Institute. In addition, *Chikyu*'s living quarters are close to luxurious compared to what researchers and crew endured on the older ship, a converted oil-exploration vessel.

Designing *Chikyu* from the hull up to be a research ship "allowed us to plan very smooth handling of the cores," says Shin'ichi Kuramoto, a seismologist with the ship's owner, the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). The layout

Ocean Goliath. The *Chikyu* is 45% longer and displaces more than twice the weight of its predecessor in the global ocean drilling program.

Cutting-edge technology. The *Chikyu* comes with a 4000-meter tube, called a riser, that encloses the drill pipe and helps it operate in difficult and unstable conditions. It works in combination with a blowout preventer (*inset*), a 300-ton device that will sit on the sea floor, to contain any explosions if the drill pokes into pressurized deposits.

Oceanography in Narragansett. He recalls installing the first microbiology lab—an aluminum storage shed of the kind seen in suburban backyards—in 1999, on selected cruises of the *Resolution*. More recently, he says, "we inherited some space [in the laboratories] and elbowed our way in." But installing key equipment, such as a radiotracer lab to determine how fast sea-floor microorganisms grow and their metabolic rates, often meant leaving somebody else's experiment behind. That could lead to some tensions on board, says Smith: "You had to basically step on someone else's toes to get this lab onboard, and then you had to go on a cruise with those people." On *Chikyu*, Smith notes, a radiotracer lab will be available on every voyage.

Long-term monitoring will also benefit from *Chikyu*'s heft. Its heavy lifting capacity will also allow researchers to place larger packages of instruments on the sea floor. Coffin says strategically placed seismometers and instruments to measure fluid flow through rock "could revolutionize our knowledge of the oceanic lithosphere."

Chikyu's designers did not forget creature comforts. Those sailing on the *JOIDES Resolution* slept in bunk beds, with up to four people in a room, and shared bathrooms and showers. In contrast, the *Chikyu* has single rooms complete with bathrooms, showers, desks, and even Internet connections for each of the roughly 50 scientists and 100 crew members expected to live on the ship for stretches of 4 to 8 weeks. There will be better recreation facilities as well.

JAMSTEC and IODP officials hope that the *Chikyu* will begin a series of shakedown cruises this fall. They are particularly interested in familiarizing the crew and scientists with the riser drilling capabilities and working the bugs out of a new on-board database system that will display on one screen all the information associated with a particular core sample.

Scientific drilling is expected to begin in earnest in the summer of 2007. The first target is the seismogenic zone of the Nankai Trough, where the Philippine Sea Plate is being forced beneath the Eurasian Plate. Achieving a better understanding of the process, which has generated some of Japan's most devastating earthquakes, presents a fitting first challenge for the world's most impressive scientific drill ship.

–DENNIS NORMILE

ensures that fragile cores will get a minimum of handling before undergoing critical testing and that biological samples will be moved quickly to oxygen-free or cryogenic storage to minimize degradation and contamination.

From the outset, *Chikyu* was designed "to go deep," says Asahiko Taira, director general of JAMSTEC's Center for Deep Earth Exploration. That translated into a 4000-meter riser, a tube that encloses the drill pipe and allows the circulation of a heavy drilling mud that lubricates the drill pipe, flushes cuttings from the drilling face, and shores up unstable sediments. The riser and a blowout preventer—a 300-ton device that sits on the sea floor—will prevent oil or gas from fouling the sea if the drill pokes into pressurized deposits.

Once the cores are extracted, they will be cut into 1.5-meter lengths and then routinely put through several nondestructive analyses never before available on a drill ship. They include a computed tomography (CT) scan, using a standard medical imager. Previously, scientists have used gamma ray scanning to image the surface of the cores. The CT scan will provide a three-dimensional image showing the porosity, microstructures, deformations, and stratigraphy of the cores' key features—data that will shed light on the geological history of the sample. The information will be used to "set a strategy for splitting the core," Kuramoto says, including selecting the best axis to expose strata or anomalies such as

hard rocks suspended in soft sediments.

Once split, core halves will go through an x-ray fluorescence (XRF) scanner. The technique is just now being introduced to earth sciences, with fewer than a dozen scanners currently available worldwide. "Right now what happens is that people take plugs at 5-cm intervals down the core, and you don't know what you've sampled until you get home and analyze it," says Boston University's Murray. XRF scanning is nondestructive and produces detailed, continuous data on the core's chemical composition. Murray says determining changes in sedimentary deposits on a millimeter scale "will lead to being able to document changes in climate at very high resolution."

Another major piece of equipment making its ship debut is a magnetically shielded chamber that blocks out 99% of Earth's magnetic field. Scientists rely on magnetic signatures locked in core samples to decipher details of plate tectonics, date sediments and rocks, and read the historical behavior of Earth's magnetic field. Previously, cores had to be taken to one of a few land-based laboratories for such measurements. JAMSTEC's Taira says soft sedimentary samples often became deformed in transit, which changed the orientation of the magnetic minerals.

Chikyu's size is a boon to microbiologists, says David Smith, a microbiologist at the University of Rhode Island's Graduate School of

Structural Genomics, Round 2

As NIH plans to extend its high-speed structural biology program for another 5 years, researchers remain divided on how to best allocate its shrinking budget

Five years ago, facing some opposition, the U.S. National Institutes of Health (NIH) in Bethesda, Maryland, launched an ambitious effort that some have compared in scale and audacity to the Human Genome Project. Its ultimate goal: to obtain the three-dimensional structures of 10,000 proteins in a decade. Like the genome project, this effort, called the Protein Structure Initiative (PSI), could transform our understanding of a vast range of basic biological processes. And just as the genome project attracted debate and dissent in its early days, the initiative split the structural biology community. The effort is now approaching a critical juncture, and the debate is heating up again.

The project is nearing the end of its pilot phase, a 5-year effort to develop technologies that has begun to transform labor-intensive, step-by-step procedures into a production-line process. Now, the initiative is poised to move into the production phase, dubbed PSI 2. In the next few months, NIH is expected to designate three to five centers, each of which could receive grants of about $12 million a year to crank out protein structures at an unprecedented clip. It will also pick a handful of smaller labs to work on problems that have so far proven difficult to solve, such as how to obtain the structures of proteins embedded in cell membranes. Officials at the National Institute of General Medical Sciences (NIGMS), which is bankrolling the initiative, are reviewing proposals for the two types of grants, and the winners are expected to be announced this summer.

But, in a debate eerily similar to the one that roiled the genome community a decade ago, structural biologists are divided on how fast to proceed—especially in the light of constraints on NIH's budget. The central issue is whether the technology is far enough along to justify the move to mass production, or whether the emphasis should continue to be on technological development.

Brian Matthews, a physicist at the University of Oregon, Eugene, and chair of PSI's external advisory board, argues that the time is ripe to move ahead in cataloging thousands of new structures. "This information will be broadly applicable to biology and medicine," he says. Raymond Stevens, a structural biologist at the Scripps Research Institute in La Jolla, California, agrees that "the technology that has come out so far has been truly impressive." But he has strong reservations about PSI 2's planned emphasis on mass-production of structures. "It's premature to start production centers until better technologies are in place," Stevens says.

This is not just an academic debate. The PSI could determine whether a key goal of structural genomics is achievable: the development of computer models to predict the structure of a new protein from its amino acid sequence. The initiative could also provide insights into how proteins interact to choreograph life's most fundamental processes and help researchers identify important new drug targets.

Picking up the pace

In one respect, the scientists who planned the human genome project had it easy. Gene sequencing relies chiefly on one technology: reading out the string of letters in DNA. By contrast, producing protein structures requires mastering nine separate technological steps: cloning the correct gene, overexpressing the gene's protein in bacteria, purifying it, coaxing it to form a crystal, screening out the best crystals, bombarding them with x-rays at a synchrotron, collecting the diffraction data as the rays bounce off the protein's atoms, and using those data to work out the protein's precise structure. (Researchers turn out a smaller number of structures using another technique known as nuclear magnetic resonance spectroscopy.)

Initially, the nine centers participating in the pilot phase of PSI had trouble dealing with that complexity (*Science*, 1 November 2002, p. 948). But structural genomics teams have now automated every step. "It took these groups a couple of years to get all the hardware in place," says Matthews. "But I think [the PSI's first phase] has been very successful."

Among the advances is a robot being built at the Joint Center for Structural Genomics (JCSG) in San Diego, California, that can run 400,000 experiments per month to find just the right conditions to coax given proteins to coalesce into high-quality crystals. Synchrotron facilities too have seen vast improvements in robotics. Setting up a crystal for measurement has historically been a cumbersome process, typically

Pure speed. Researchers at the Midwest Center for Structural Genomics use robotic gear to speed protein purification.

taking hours of fine-tuning. JCSG researchers and others have now created robotic systems to carry out this work, enabling data collection on up to 96 crystals without interruption. "That has been a tremendous benefit," says JCSG chief Ian Wilson, a structural biologist at the Scripps Research Institute.

As the technologies advanced the centers accelerated their output. They produced 350 structures in PSI's fourth full year, up from just 77 in the first year, and are on track to complete 500 this year. That pace is still well short of the initial goal of 10,000 structures in 10 years—that goal was little more than an optimistic guess, PSI leaders now say—but it's a big step forward and should be fast enough to accomplish most of the effort's scientific goals. Equally important, says John Norvell, who directs NIGMS's PSI program, the average cost of each structure has dropped dramatically, from $670,000 in the first year—a number inflated by the cost of purchasing and installing robotic gear—to $180,000 in year 4. This year, Norvell expects that the cost will drop to about $100,000 per structure. By comparison, he adds, traditional structure biology groups typically spend $250,000 to $300,000 for a structure, although some of the proteins they tackle are far more complicated than those PSI has taken on.

The types of proteins targeted by PSI are, however, one bone of contention. Traditional structural biology groups tend to go after similar proteins in important families, such as kinases, that participate in many biological pathways. And they often determine the structure of complexes of one protein bound to different molecular targets, in order to tease out the details of how the protein functions. As a result, 87% of the structures deposited in the major global protein database are closely related to those of other proteins.

The PSI, however, was set up to acquire structures from as many of the estimated 40,000 different protein families as possible. Indeed, 73% of the structures the PSI centers have solved so far have been "unique," which by the PSI definition means that at least 30% of the gene sequence encoding a protein does not match that encoding any other protein. The idea behind casting such a broad net is to acquire structures from representatives of each family in the hope that this will enable computer modelers to predict the structure of other family members. Already, the data suggest that there is not as much

A Dearth of New Folds

The Protein Structure Initiative (PSI) has already come up with one surprise: Proteins apparently come in a relatively limited variety of shapes. The initiative is targeting "unique" proteins, ones in which the DNA that encodes them differs markedly from that for proteins

Protein landscape. This graph reveals how proteins cluster into four structural classes.

with a known structure. Researchers expected that many if not most of those proteins would have structural patterns never seen before, but the vast majority look quite familiar.

The general shape of a protein once it assumes its three-dimensional (3D) form is known as a fold. So far PSI groups have found that only 12% of their completed structures sport new folds. "The number of folds will be considerably less than previously thought," says Ian Wilson, a structural biologist at the Scripps Research Institute in La Jolla, California, and head of its Joint Center for Structural Genomics. This means that proteins with vastly different patterns of amino acids adopt similar 3D shapes. That, Wilson says, is critical information for computer modelers working to predict the structures of proteins based only on their DNA sequence.

Researchers are also mapping out how all these unique proteins relate to one another. In a report published online on 10 February by the *Proceedings of the National Academy of Sciences*, researchers at the University of California, Berkeley, and the Lawrence Berkeley National Laboratory (LBNL) in California compared nearly 2000 different protein structures, calculating the difference in shape between each protein and all of the others in the collection. They then graphed the results, showing similar structures as close to one another. They found that the global protein structure landscape is a bit like the cosmos, where galaxies cluster together amid vast regions of emptiness.

That map does have sharp features, however, says study author Sung-Hou Kim, an LBNL structural biologist and the head of the Berkeley Structural Genomics Center. It shows the four main classes of protein structures—known as α helixes, β strands, and proteins with mixtures of α and β domains called $\alpha+\beta$ and α/β—as four elongated arms emerging from a common center. The map, Kim says, suggests that much of the protein structure space is empty because proteins with certain shapes are architecturally unstable. That in turn suggests that structural genomics groups are unlikely to find any new structural classes of proteins. Says Kim: "I would be very surprised if they did." **–R.F.S.**

structural variation between families as many biologists expected (see sidebar).

Some structural biologists argue, however, that this approach has limited value, and that the tens of millions of dollars currently going to structural genomics centers would be better spent on traditional structural biology groups. Yale biochemist Thomas Steitz, for example, says that most of the structures PSI groups have produced so far are "irrelevant" to understanding how the proteins work because they are not

bound to their targets. The PSI focuses on bacterial rather than eukaryotic proteins, he also complains.

Berg acknowledges that "tension certainly exists," between traditional structural biologists and structural genomics groups. Although some PSI 2 centers will likely focus on producing structures of protein complexes and eukaryotic proteins, he notes that NIH's structural genomics effort was never set up to go after the same type of information as conventional structural biol-

Numbers game. This protein crystallization robot aids structure solving.

ogy. Rather, the goal was to explore the far reaches of the protein landscape. "To my mind the most important message is structural genomics and structural biology are largely complementary and synergistic," Berg says.

Berg and others add that the PSI has already provided numerous important biological insights. For example, the Northeast Structural Genomics Consortium (NESGC) recently solved the structure of a protein that adds a methyl group to ribosomal RNA and in the process confers antibiotic resistance to bacteria. That structure, says NESGC director Guy Montelione, has suggested inhibitory compounds that could revive current antibiotics and spawned a separate research program on the topic. Another structure revealed details of the way plants bind a signaling molecule called salicylic acid, challenging conventional wisdom on the functioning of plants' immune systems. "Not only are we spinning out new science, but new science initiatives," Montelione says.

Chapter 2

What comes next is, however, a matter of debate. NIGMS officials had expected to scale up a handful of the current PSI centers to full-scale production facilities and fund as many as six additional technology centers, each tackling a separate bottleneck.

A tight NIH budget has already forced NIGMS officials to rethink those plans, however. They had hoped to boost the current PSI budget of $68 million to $75 million next year, the first year of PSI 2. But they are now anticipating a decline in funding, to $64.5 million. The cut is likely to force structural genomics leaders to rein in their goals, and ultimately it could extend the date by which they complete the program. "That's clearly going to be a problem," Matthews says.

Those budget cuts will also make it tough for PSI leaders to strike the proper balance between production and technology development in the next phase. Each of the existing pilot centers currently receives some $8 million a year. The plan for PSI 2, Norvell says, had been to spend $12 million a year on each production center. That means five centers would eat up nearly the entire budget for PSI 2's first year, leaving little for technology development. If that happens, "I think we'll regret it in 5 years," Stevens says.

Stevens points out that many technical problems remain. For
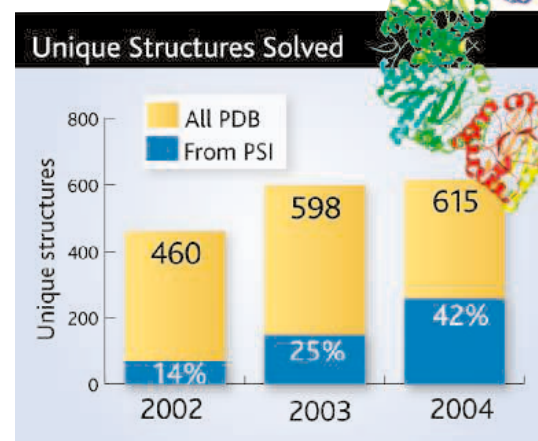
example, even though PSI centers have increased their output of protein structures, their success rate in turning targeted genes into solved structures has remained essentially unchanged. At each stage in determining a structure—cloning the gene, expressing the protein, and so on—researchers take a hit. For example, only 57% of cloned genes are successfully expressed as proteins, and of those, only 28% can be purified. "It's like doing chemical synthesis" that involves numerous steps, says Wilson. "If you have a 90% success rate at each step, that's not going to give you much material out at the end."

In the end, Stevens notes, only 2% to 10% of the proteins targeted by PSI centers wind up as solved structures. In view of this "pretty poor success rate," Stevens argues that the phase 2 efforts should focus more on technology development. "I think structural genomics can do even better if the technologies are allowed to mature further," he says.

Not many of Stevens's colleagues agree. "Clearly we have to capitalize on the production centers we've already invested in," says Wilson. Thomas Terwilliger, a structural biologist at the Los Alamos National Laboratory in New Mexico and head of the TB Structural Genomics Consortium, adds that the limited success rate isn't a major issue because if one protein in a family doesn't yield a structure, researchers can typically find another one that does. Furthermore, Montelione points out that the new production centers will spend about one-third of their funds on improving the technology. Stevens counters that "there will be so much pressure to produce structures that any technology developments will take a significant back seat to the structure focus."

Berg says "it's hard to imagine funding fewer than three of the large-scale [production] centers." At $12 million apiece, that would still leave $28.5 million—more than $4 million for each of the six proposed technology centers. The balance between production and technology development is "still very much up in the air," says Berg, and will depend on the outcome of the reviews of the grant proposals. The NIGMS advisory committee will then decide which centers to fund in May and announce their decision in early July.

Whatever the outcome, it's now unlikely that the PSI effort will achieve the initial goal of solving 10,000 protein structures by 2010. With budget cutbacks and continued technical challenges, the final tally will probably be somewhere between 4000 and 6000, about the number that PSI leaders now believe computer modelers will need to accurately predict structures of related family members. Still, that means the program will solve structures for only a small fraction of the estimated 40,000 protein families. "This mixed bag of production and technology development will require another cycle, another 5 years to finish the job," says Montelione. So will there be a PSI 3? That debate is just starting.

–ROBERT SERVICE

RANDOM SAMPLES

Edited by Constance Holden

Non-Euclidean Potholders

Knot theory is no longer the only branch of mathematics that appeals to the handicrafts set. A big crowd showed up last month at the Kitchen, one of Manhattan's hippest performance spaces, to hear a pair of Cornell mathematicians talk about hyperbolic space. Their main props: crocheted models of objects in the hyperbolic plane, a central concept in non-Euclidean geometry. In ordinary Euclidean space, a flat plane stretches out forever and parallel lines never meet, explained geometers Daina Taimina and David Henderson. However, as mathematicians discovered in the early 1800s, that's not true in other kinds of space. In the sphere, for example, parallel lines meet at the poles. In the hyperbolic plane, which can be thought of as the opposite of the sphere, parallel lines shy away from each other.
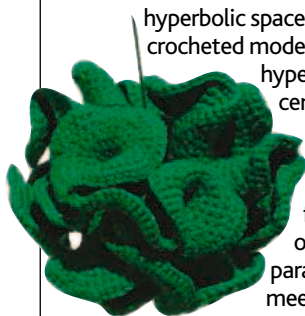
Hyperbolic "pseudosphere."

Hyperbolic space is very hard to picture. For more than a century, mathematicians struggled without notable success to make 3D models of it. Then Taimina had the idea of using her crochet hook. She and Henderson use crocheted models in their classes at Cornell and hope that when people create and play with the objects—which look like witches' hats, flamenco skirts, or curly kale—they'll develop an intuitive sense of what hyperbolic geometry is all about. "We all play with balls as children. With a sphere, you have that memory in your hands. But you don't have that with hyperbolic geometry," says Taimina.

Man the Eroder

For the past millennium, humans have been moving more earth than all natural processes combined. Just how far have we tipped the balance? Geologist Bruce Wilkinson of the University of Michigan, Ann Arbor, decided to find out.

He calculated prehistoric rates of erosion through the amount of sedimentary rock, the end result of erosion, that has accumulated over the past 500 million years and estimated that natural erosion lowers Earth's land surface about 24 meters every million years. He then calculated the human contribution, combining estimates of erosion from crop tillage, land conversion for grazing, and construction. Averaged out over the world's land surface, that came to about 360 meters per million years, or 15 times the natural rate.

This difference amply demonstrates that current agricultural practices are unsustainable, says Wilkinson, who points out in this month's issue of *Geology* that at the current rate, the soil eroded from Earth's surface would fill the Grand Canyon in 50 years.

Wilkinson's estimates for natural erosion are similar to those of geologist Paul Bierman of the University of Vermont, Burlington, who has used beryllium isotopes to estimate erosion rates in the Appalachian Mountains from the past 10,000 to 100,000 years. "The most intriguing part of this study is to be able to look back over 500 million years of earth history," says Bierman.

Life in the Air

Gene sequencer extraordinaire J. Craig Venter has launched yet another bold venture: inventorying the DNA from bacteria, viruses, fungi, and other microbes floating around in the air.

Venter is currently engaged in a round-the-world project cataloging the organisms—and in particular, their genes—in seawater following a pilot project in the Sargasso Sea (*Science*, 2 April 2004, p. 66). He has chosen Manhattan as the test bed for a new Air Genome Project.

Using $2.5 million from the New York City–based Alfred P. Sloan Foundation, the J. Craig Venter Institute in Rockville, Maryland, has begun analyzing the material collected from a rooftop filter in midtown Manhattan. There's been "a lot of basic groundwork in terms of designing and trying out different air-sampling devices," says Venter. The U.S. Department of Homeland Security does some air biomonitoring, he notes, but only for a few hazardous things like anthrax: "Nobody has any idea what the background in the atmosphere is."

Venter hopes the project will go beyond antibioterrorism to finding out "who is there," including organisms that affect health. He's still tinkering with the technology, which builds on his pioneering use of the whole-genome shotgun approach to explore undefined populations. The institute also plans to collect samples inside buildings. All the data will be put in the public domain.
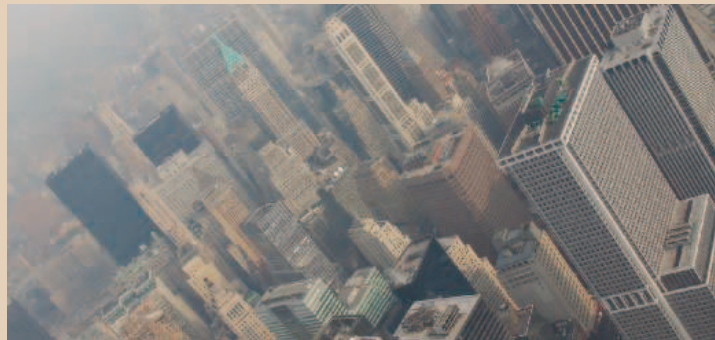
What Manhattanites breathe.

Edited by Yudhijit Bhattacharjee

NONPROFIT WORLD

Building bridges. After 30 years in England, U.S.-born anthropologist Leslie Aiello is coming home. Next month, Aiello, a human evolution expert at University College London (UCL),

becomes president of the Wenner-Gren Foundation for Anthropological Research in New York City. Last year, the foundation gave out 190 grants totaling over $3 million.

Aiello has spent nearly all of her career at UCL, gradually moving from research into administration. "I found that I liked to make things happen," she says.

At Wenner-Gren, Aiello, 58, hopes to bridge the gap between biological and social anthropologists. "If we don't keep anthropology as a unified discipline," she says, "we are in danger of losing some of the spark that could lead to major advances."

AWARDS

First choice. India's new science adviser is the inaugural winner of the country's biggest scientific prize. But officials say his high rank in government has nothing to do with his being chosen for the honor.

Chemist C. N. R. Rao, who was appointed chair of the Scientific Advisory Council to the Prime Minister in January, wins the $62,500 Indian Science Award for his contributions to solid state chemistry and materials science. A government-appointed panel of 12 scientists

from India and overseas picked Rao, 71, in a process that began in early 2004, several months before the current government came to power. "We cannot disqualify a person for being in a certain position,"

says India's science minister Kapil Sibal. "The award goes to him for his excellent past work."

The work has earned recognition outside India, too: Last month, Rao shared the $1 million Dan David Prize (see below) in the Future Time Dimension with Harvard chemist George Whitesides and MIT chemical engineer Robert Langer.

■ Archaeologists Graeme Barker and Israel Finkelstein are the joint winners of the $1 million Dan David Prize in the Past

Time Dimension. Barker, a professor at Cambridge University in the U.K., is honored for his contributions to landscape and environmental archaeology, while Finkelstein, a professor at Tel Aviv University in Israel, is recognized for applying archaeological knowledge to reconstruct biblical history. British theater director and filmmaker Peter Brook was honored for the Present Time Dimension. The prize is awarded by the Dan David Foundation and Tel Aviv University.

Open-ended inquiry. Physicist Charles Townes, who received the Nobel Prize in 1964 for inventing the maser and co-inventing the laser, has won the $1.5 million Templeton Prize for his efforts to bridge science and spirituality. The annual award, from the John Templeton Foundation, recognizes individuals who have "advanced knowledge in spiritual matters."

Two years after winning the Nobel, Townes generated controversy with an article on the convergence of science and religion. He has continued to write and talk about the subject. "It is important for us to be open-minded in science and religion. The two are more similar than one may think," he says.

Townes, 89, plans to donate half of the prize money to his alma mater, Furman University in Greenville, South Carolina. Much of the rest will go to faith-based institutions.

THE EXTRA MILE

Deferred honor. A winner of Germany's top scientific award won't accept a $2 million prize until her university completes an investigation she requested of a paper from her lab.

Stefanie Dimmeler, a cardiologist at Wolfgang Goethe University in Frankfurt am Main, is one of 10 winners of the Leibniz prize, awarded this month by the German research agency, DFG. Her celebrity revived charges raised a year ago that a figure Dimmeler and her colleagues published in *Nature Medicine* in November 2003 was identical to one published a few months earlier in *Blood*. The mistake was the result of a confusing computer system for storing images, Dimmeler says. She and her colleagues repeated the experiments, reached similar conclusions, and published corrections in both journals.

Dimmeler notified the DFG when the mistake was discovered, but the agency decided then that there was no need to investigate. When the Leibniz winners were announced in December, however, an anonymous letter to the DFG raised the issue again and suggested that Dimmeler didn't deserve the prize. To clear up any remaining questions, Dimmeler and the DFG asked her university to conduct an investigation. "We agreed that we should do everything in the most correct way," she says. The panel is expected to finish its work by May.

Got any tips for this page? E-mail people@aaas.org

LETTERS

Academy of Natural Sciences: Job Cuts

JOCELYN KAISER'S ARTICLE ON JOB CUTS AT THE Academy of Natural Sciences in Philadelphia ("Philadelphia institution forced to cut curators," News of the Week, 7 Jan., p. 28) exemplifies a disturbing trend that threatens our understanding of biological diversity. At a time when species are thought to be going extinct at record rates (*1*), our capacity to describe that diversity is being severely undermined. The situation in Philadelphia demonstrates that the very institutions charged with this cause are now also being threatened with extinction.

Academy President D. James Baker does not seem to understand this, and his vision for the institution is a frightening prospect for the entire natural history museum community. Efforts to focus Academy research on noncollections-based programs such as watershed management are misdirected. Such programs already exist at universities and environmental consulting firms around the country, and reproducing them devalues the very thing that makes the Academy unique—its biological collections. The Academy is a taxonomic institution and that should remain its central focus. The history of the Academy suggests that, once a curator is lost, the associated collection falls into obscurity, and now ornithology at the Academy is threatened. Furthermore, Baker's implication that a taxonomic focus cannot bring in outside research dollars is a fallacy. At the same time, systematists cannot be expected to bear the burden of fixing their institution's financial situation. What the Academy needs is enthusiastic leadership that understands its institution's taxonomic mission. Baker and the Academy board seem to lack this understanding.

JOHN S. LAPOLLA

Department of Entomology, Smithsonian Institution, Post Office Box 37012, NHB, CE518, MRC 188, Washington, DC 20013–7012, USA.

Reference
1. S. L. Pimm, P. Raven, *Nature* **403**, 843 (2000).

THE ACADEMY OF NATURAL SCIENCES IN Philadelphia is one of the most important research museums in the world, with a rich tradition going back to Audubon and beyond. Unfortunately, its stature is now in grave jeopardy because of cuts in staff ("Philadelphia institution forced to cut curators," J. Kaiser, News of the Week, 7 Jan., p. 28).

Reductions in Academy staff were necessary because of a severe budget deficit, but the nature of the cuts signifies a major problem in leadership at the institution. In scrambling for dollars, the Academy's directors have lost sight of the institution's fundamental mission. The budget cuts disproportionately slashed basic museum research. For example, the bird collection must now operate without a research head for the first time in almost 200 years. What remains after the cuts are mostly cash-cows, namely, exhibits and applied research in environmental and biomedical science. Although exhibits are important to the Academy's mission, applied programs are not. The Academy is a natural history museum, not the Environmental Protection Agency or the Centers for Disease Control.

Heads should roll at the Academy, that is clear, but not the heads of employees who are fulfilling the mission of the institution.

FREDERICK H. SHELDON, J. V. REMSEN, ROBB T. BRUMFIELD

Museum of Natural Science, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, USA.

IN HER NEWS OF THE WEEK ARTICLE "Philadelphia institution forced to cut curators" (7 Jan., p. 28), J. Kaiser reports on budget shortfalls at the Academy of Natural Sciences in Philadelphia. In my discussions with Kaiser, I emphasized that the cutbacks mentioned in the article were made to reduce the Academy's budget deficit and should not be construed as reflecting negatively on the individuals concerned. It was unfortunate that the article gave the names of the curators who received notice and implied that these individuals were laid off because of unsatisfactory performance.

D. JAMES BAKER

President and CEO, Academy of Natural Sciences, 1900 Benjamin Franklin Parkway, Philadelphia, PA 19103, USA.

The Recreational Fisher's Perspective

IN THEIR REPORT "THE IMPACT OF UNITED States recreational fishers on marine fish populations" (Reports, published online 26 Aug. 2004, DOI 10.1126/science.1100397; 24 Sept. 2004, p. 1958), F. C. Coleman *et al.* suggest that the overfished condition of marine fish stocks rests on the shoulders of the recreational angler. In reality, the study merely confirms what fishery scientists,

managers, and anglers themselves have known for decades—recreational catches comprise a significant share of some U.S. fisheries. Unfortunately, the authors have mischaracterized marine fisheries by pointing out just the current proportion of recreational landings of a few "populations of concern." What the authors fail to consider in the study is how most of these species arrived at their current condition—through years of commercial overfishing.

A boccacio

For example, the authors say that in 2002, anglers landed 87% of the total harvest of bocaccio, or approximately 200,000 pounds. What the authors fail to explain is that of the total 104 million pounds of bocaccio landed in the previous 20 years, commercial fisherman landed 89 million pounds, more than 85% of the total. In other words, over the same period, recreational landings account for less than 15% of the total (*1*). It is the sustained commercial overfishing of bocaccio (see figure in Supporting Online Material) (*2*) that is the primary reason for driving this species into decline.

The authors also fail to accurately represent the reality of the Gulf of Mexico red snapper fishery. To state that recreational anglers take half of the total red snapper harvest is to ignore the most significant part of the story. The authors never mention the source of mortality that has the greatest impact on red snapper stock recovery: mortality of juvenile snapper caused by shrimp trawl bycatch. The reality is that even if all sportfishing—and commercial fishing, for that matter—ended today, the stock will never recover without addressing this major source of mortality.

CREDIT: MARK CONLIN PHOTOGRAPHY

In the United States, saltwater recreational anglers are highly regulated by state, interstate, and federal bag limits; size limits; and seasons, the same tools that have successfully managed freshwater anglers for many years. The anglers, and the industry they support, have a strong tradition of supporting and paying for good fisheries management. This study does nothing more than malign this community and mislead the American public at a time when we all need to be working cooperatively to rebuild our fisheries.

MICHAEL NUSSMAN*

President, American Sportfishing Association, 225 Reinekers Lane, Suite 420, Alexandria, VA 22314, USA.

*On behalf of the American Sportfishing Association, American Fly Fishing Trade Association, B.A.S.S., Inc., The Billfish Foundation, Canadian Sportfishing Industry Association, Coastal Conservation Association, Coastside Fishing Club, Congressional Sportsmen's Foundation, Federation of Fly Fishers, International Association of Fish and Wildlife Agencies, The Izaak Walton League of America, Inc., Jersey Coast Anglers Association, National Marine Manufacturers Association, Northwest Sportfish Industry Association, Recreational Fishing Alliance, Sportfishing Association of California, Stripers Forever, United Anglers of California, and United Anglers of Southern California

References

1. A. D. MacCall, "Status of bocaccio off California in 2003" (Santa Cruz Laboratory, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Santa Cruz, CA, June 2003).
2. See Supporting Online Material on *Science* Online at www.sciencemag.org/cgi/content/full/307/5715/1560c/DC1.

Response

IT IS OUR VIEW, BASED ON THE DATA, THAT THE same tools used to manage freshwater anglers have not proved effective in constraining the cumulative effect of saltwater recreational fishing, nor has the entire toolbox used in freshwater management been applied in saltwater, including lotteries and annual bag limits. Such facts point to the need for better (not more) regulations that effectively stop overfishing so that both recreational fishing and commercial fishing are sustainable enterprises.

Nussman contends that we "suggest that the overfished condition of marine fish stocks rests on the shoulders of the recreational angler." To the contrary, we simply point out that recreational fishing takes 23% of these overfished stocks, based on U.S. National Marine Fisheries Service (NMFS) data and on the NMFS's most recent status report on U.S. fisheries (*1*). Clearly, commercial fishing plays a significant role in taking the remaining 77%. Further, commercial fishing has played a significant role in the catch histories of individual species.

Many, indeed most, of the overfished stocks can be attributed to commercial fishing. However, this is not the case for all species. In the Gulf of Mexico, four of the five most productive species that are overfished are taken primarily by recreational anglers and have been for over most of the past 22 years. (On the Atlantic coast, bluefish catch has steadily declined to 18% of 1981 levels and has been primarily recreational, while black sea bass shifts back and forth between the two.)

Nussman states that we inaccurately depict the red snapper fishery by not addressing bycatch. The directed red snapper fishery typically points to bycatch of juvenile red snapper as the single most important factor inhibiting recovery, while the shrimp fishery suggests that the directed fishery is to blame for removing the largest, oldest, and most fecund individuals, thus truncating the age and size structure of the population. This argument has persisted now for decades, but could be more clearly resolved by improving the poor estimates of natural mortality in both juvenile and adult stages. Recent stock reduction analyses (*2*, *3*) suggest that shrimp trawl bycatch of juvenile red snapper has had relatively little impact on the depletion of the red snapper stock, and indicate instead that the mortality rates from commercial and recreational fishing have caused the large depletions in stock abundance.

Our objective in conducting this body of work was to inform the public that both commercial and recreational sectors contribute to overfishing. The ecological and economic sustainability of these sectors depends on acceptance of this shared responsibility; knowledge of all sources of mortality (including agricultural pollution, industrial pollution, and coastal development); and cooperation to rebuild healthy populations and ecosystems.

FELICIA C. COLEMAN,[1] WILL F. FIGUEIRA,[2] JEFFREY S. UELAND,[3] LARRY B. CROWDER[4]

[1]Department of Biological Science, Florida State University, Tallahassee, FL 32306–1100, USA. [2]Department of Environmental Sciences, University of Technology Sydney, Westbourne Street, Gore Hill, NSW 2065, Australia. [3]Department of Geography, Clippinger Lab 122, Ohio University, Athens, OH 45701, USA. [4]Nicholas School of the Environment and Earth Sciences, Duke University, 135 Duke Marine Lab Road, Beaufort, NC 28516–9721, USA.

References

1. U.S. National Marine Fisheries Service, Annual Report to Congress on the Status of U. S. Fisheries–2003 (U. S. Department of Commerce, NOAA, National Marine Fisheries Service, Silver Spring, MD, 2004) (available at www.nmfs.noaa.gov/sfa/reports.html#sos)
2. M. McAllister, personal communication.
3. C. Walters, personal communication.

Global Impact of Recreational Fisheries

F. C. COLEMAN *ET AL.*'S ANALYSIS OF THE IMPACT of recreational fisheries on marine fish populations in the United States ("The impact of United States recreational fisheries on marine fish populations," Reports, published online, 26 Aug. 2004; DOI 10.1126/science.1100397; 24 Sept. 2004, p. 1958) constitutes a timely contribution about the potential biological importance of this fisheries sector. However, the evidence presented is only a cursory acknowledgement of the global impact of recreational fisheries.

The authors present time series of total harvest and percentages of landings (i.e., the share that recreational anglers have of the total harvest in the marine environment as compared with commercial fisheries), which illustrates that recreational fisheries can and do sometimes contribute substantially to total harvest, particularly among some top predators of the food webs. However, the recognition that some marine species were primarily harvested by recreational anglers does not explicitly suggest that recreational fishing can have "serious ecological and economic consequences on fished populations," as Coleman *et al.* state. It is not possible to derive conclusions about ecological impacts on the basis of harvest ratios alone, let alone the economic perspective. Although we "believe" that the author's contention may be correct, it would be more appropriate to restrict their conclusions to those clearly supported and warranted by the data.

In this way, it is clear that the total harvest of recreational fisheries is decreasing over time in all stocks except in the Gulf of Mexico (their fig. 1C) and that the general share on total landings has been stable since the 1990s (fig. 1A). Does this mean that the impact of recreational fisheries is decreasing, or at least not increasing? We simply do not know yet.

There is another issue that needs to be addressed in the future. In all review papers recently published on the future of the world's fisheries [e.g., (*1*, *2*)], only commercial marine fisheries have been considered, whereas inland fisheries have not been accounted for appropriately. However, in inland fisheries of all industrialized societies, commercial fisheries have largely been replaced by recreational fisheries (*3*, *4*). Therefore, harvest ratio studies (e.g., Coleman *et al.*) would provide little insight, as fishing is conducted primarily recreationally. But does this indicate that the impact of recreational fishing is negligible in freshwater ecosystems? As much of the

world's ichthyofaunal diversity is confined to the freshwater environment, intensive, typically highly selective recreational fishing and discard mortality [e.g., (5)] coupled with deleterious recreational fisheries management actions such as harmful stocking practices (3, 6) may often lead to much stronger negative ecological and possibly evolutionary impacts of leisure fisheries in freshwater. Although the Coleman et al. paper is focused on marine fisheries in a specific jurisdiction, we contend that this type of analysis is required on a more global and inclusive scale, incorporating inland fisheries. A recent article (7) supports Coleman et al.'s position that global fisheries impacts are indeed greater than previously assumed if recreational fisheries are considered. Intensified and long-term research efforts are needed in all aquatic environments to answer this and other questions surrounding recreational fisheries, to improve fisheries management and conservation, and to move fisheries towards sustainability on a global scale.

ROBERT ARLINGHAUS,[1]* STEVEN J. COOKE[2]

[1]Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Department of Biology and Ecology of Fishes, Müggelseedamm 310, 12587 Berlin, Germany. [2]Centre for Applied Conservation Research, Department of Forest Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

*To whom correspondence should be addressed. Email: arlinghaus@igb-berlin.de

References

1. D. Pauly et al., Science **302**, 1359 (2003).
2. R. Hilborn et al., Annu. Rev. Environ. Res. **28**, 359 (2003).
3. R. Arlinghaus, T. Mehner, I. G. Cowx, Fish Fish. **3**, 261 (2002).
4. R. Arlinghaus, Berichte IGB **18**, 1 (2004).
5. J. R. Post et al., Fisheries **27** (no. 1), 6 (2002).
6. G. K. Meffe, Conserv. Biol. **6**, 350 (1992).
7. S. J. Cooke, I. G. Cowx, BioScience **54**, 857 (2004).

Response

ARLINGHAUS AND COOKE FIND OUR PAPER meaningful in that it points to a need for a more critical examination of recreational fisheries management. Their main objections seem to address issues outside the scope of our paper. We never intended to offer a global view of recreational fisheries, freshwater fishing, or stock enhancement. Instead, we presented a time series of landings in metric tons (not percentages) of all federally managed stocks in the continental United States and then presented regional landings of only those stocks considered overfished or experiencing overfishing. We did not include in the regional assessment any species not falling into this category, whether primarily taken by recreational fishers or not. The fact that these species are already considered overfished by the U.S. government implies that they are suffering serious ecological and economic consequences of intense fishing pressure. That many of those species are primarily taken by recreational fishermen does not in and of itself fault the recreational fishery sector. In some cases, the recreational percentage of landings rises primarily as a result of commercial fishery declines; this is particularly obvious in the northeast and Pacific. But in the Gulf of Mexico and the South Atlantic, the recreational component has been consistently large over the last 20 years, and in the Pacific, the recreational landings doubled between 2001 and 2002.

Our point is that scientists and managers must develop methods to constrain exploitation, whether commercial or recreational, if they are to achieve the societal goal of sustainable fisheries. If this is not the goal, then laissez les bons temps rouler.

FELICIA C. COLEMAN,[1] WILL F. FIGUEIRA,[2] JEFFREY S. UELAND,[3] LARRY B. CROWDER[4]

[1]Department of Biological Science, Florida State University, Tallahassee, FL 32306–1100, USA. [2]Department of Environmental Sciences, University of Technology Sydney, Westbourne Street, Gore Hill, NSW 2065, Australia.

[3]Department of Geography, Clippinger Lab 122, Ohio University, Athens, OH 45701, USA. [4]Nicholas School of the Environment and Earth Sciences, Duke University, 135 Duke Marine Lab Road, Beaufort, NC 28516–9721, USA.

The Discoverers of Glass

I FOUND THE ESSAY "A WORLD OF GLASS" BY A. Macfarlane and G. Martin (3 Sept. 2004, p. 1407) interesting and informative. However, the section titled "A Brief History of Glass" perpetuated a common piece of misinformation, namely, the origin of glass-making. They stated that "[glass] may have appeared first in the Middle East and regions such as Egypt and Mesopotamia around 3000 to 2000 B.C. … Glass was almost certainly discovered by accident… by Phoenician traders, who apparently noticed that a clear liquid formed when the nitrate blocks on which they placed their cooking pots melted and mixed with sand from the beach."

In his book *The Glass Makers*, Samuel Kurinsky (*1*) establishes that the early Hebrews were the first to discover the art of glassmaking from raw materials and that they maintained that secret exclusively for an extended period of time. This was recognized by many rulers from Egypt to Rome for 3000 years. The Hebrew contribution to the advancement of civilization in developing the ability to make glass from raw material was not acknowledged by the Essay authors.

SANDER J. BREINER

Michigan State University and Wayne State University, Franklin Pointe Office Center, 7457 Franklin Road, Suite #304, Bloomfield Hills, MI 48301–3604, USA.

Reference
1. S. Kurinsky, *The Glass Makers: An Odyssey of the Jews* (Hippocrene Books, New York, 1991).

A New Climate Research Center in Italy

THE NEWS ITEM ABOUT THE ITALIAN CLIMATE Research Center that will be located at the University of Lecce in Bologna ("Italy hosts a climate research center," ScienceScope, 24 Dec. 2004, p. 2171) unfortunately does not tell the whole story. The center was established without significant input from the Italian climate research community. A call for proposals was issued in 2001, and two groups responded. The review process that awarded the center to Lecce was never explained. The Ministry of the Environment simply made an informal announcement that the appropriation would go to the National Institute of Geophysics and Vulcanology, which will coordinate the center. At the moment, nothing is known about the center's programs, whether it will open to the larger climate research community, or how these programs will be funded. The initial program presented by the director suggests that the Center will be limited to oceanographers and marine ecosystem researchers in the Bologna area.

Italy has allowed its National Climate Research Program to languish, and unless significant changes are made in the scope and mission of the Climate Research Center, it will be useless. First, it seems designed to produce a climate model that closely resembles those already developed elsewhere, reminiscent of Wittgenstein's philosophy that it is better to buy several copies of the same newspaper to be nearest the truth. Second, without the involvement of the larger climate academic community, the program will have no impact on higher education in the rest of the country.

GUIDO VISCONTI

Center of Excellence for Remote Sensing and Severe Weather, University of L'Aquila, Via Vetoio, 67010 Coppito, L'Aquila, Italy.

A Blue View of the Red Planet

Victoria E. Hamilton

In *Mars: A Warmer, Wetter Planet*, Jeff Kargel does an admirable job of tying together the work of many scientists whose results provide evidence for the hypothesis that Mars may have periodically had a warmer, wetter climate than previously recognized. Kargel, an astrogeologist with the U.S. Geological Survey in Flagstaff, Arizona, integrates and explains such diverse work as geomorphologic observations, theoretical calculations, and measurements of mineralogy and chemistry obtained from orbiting spacecraft. The book is full of excellent examples of the most recent imaging data from NASA missions to Mars, as well as many photos of terrestrial analogues, which surely will aid the nonspecialist in understanding how planetary geologists go about understanding the landforms of Mars. The book emphasizes image-based comparative geomorphology, but Kargel also devotes text and many of the color plates to non-imaging data such as compositional, topographic, and slope maps as well as spectral analyses and other quantitative information.

> **Mars**
> **A Warmer,**
> **Wetter Planet**
> *by Jeffrey S. Kargel*
>
> Springer, New York, and Praxis, Chichester, UK, 2004. 603 pp. Paper, $34.95, £20. ISBN 1-85233-568-8.

My favorite of the color plates is the first, a copy of Hugh Kieffer's hand-drawn plot of the relationships among data (numbers), information (numbers assembled in some meaningful order), perceived knowledge (what we think we know), and understanding (what we actually know). The graph reminds us that, despite the ever-increasing quantity of data and information, human knowledge increases slowly (with ups and downs) and that true understanding lags even farther behind. I first encountered the diagram as a graduate student, when Kieffer drew it at a meeting of the Mars Global Surveyor Thermal Emission Spectrometer science team. It impressed me as it has the author, and I am happy to see it preserved for a wider audience, because it applies to virtually all human endeavors

The reviewer is at the Hawaii Institute of Geophysics and Planetology, University of Hawaii, 2525 Correa Road, Honolulu, HI 96822, USA. E-mail: hamilton@higp.hawaii.edu

that seek to learn more about how nature works. In addition to presenting such lessons and the geologic story of water on Mars, Kargel covers in reasonably technical detail the highly topical (and controversial) issues of the evidence for life in a martian meteorite, the potential for life on Mars, future unmanned and manned exploration of the planet, permanent human habitation and requisite mining of natural resources, and the terraforming of Mars.

In person, Kargel comes across as a person very passionate about planetary science—and Mars in particular—but also as a keen observer of history and his fellow inhabitants on Earth. His writing leaves the same impression: He places the exploration of Mars in context through references to concurrent historical events (and song lyrics) and peppers his story with personal philosophical ruminations. Although many readers will appreciate and enjoy the historical connections, some of the examples (such as the flaming Cuyahoga River or an oblique reference to the Chappaquiddick scandal) might be lost on those under the age of 35 or who were not born and raised in the United States.

The breadth of subject material, level of technical detail, and goodly doses of the author's philosophical musings make the book a rather lengthy read. The publisher describes the intended audience of the book as "amateur astronomers and space science enthusiasts." In truth, readers who lack a reasonable working knowledge of

Play lava or platy mudflow?
The broken platy flow features of Marte Vallis were formed by a very fluid material that had a thin surficial crust.

basic geology, glaciology, and geochemistry—or easy access to a good geologic dictionary—may have some difficulty fully comprehending various illustrations and descriptions in Kargel's scientifically dense and jargon-packed account. On the flip side, specialists and enthusiasts will find that the inclusion of detailed information drawn from a wide range of subdisciplines makes for a well-rounded story.

Kargel emphasizes the utility of multiple working hypotheses and does not shy away from presenting alternative views. Nonetheless, he does not always thoroughly describe such alternatives, and he almost always concludes that water in some form was responsible for the features he describes. The reader cannot help but be left with the overall impression that—despite the mentioned alternatives and the admonitions that unusual or unknown martian conditions and processes be considered—in the end, the author favors a water-related interpretation every time. Nonspecialist readers may not realize that many specialists will disagree with some of these interpretations. In fact, I can envision the book being used in graduate-level planetary science seminars that seek to critically evaluate the evidence for the role of water and water ice in shaping the present surface of Mars. As with all scientific inquiry, various hypotheses will be supported, refuted, or retooled based on our interpretations of additional information. In the meantime, Kargel's entertaining book presents one evolving path of scientific thought regarding Mars that is gaining wider acceptance. It will instigate useful, stimulating debate on the role of water in shaping the Mars we see today and point to how we might enhance our understanding of the Red Planet in the future.

EXHIBITS: EXPLORATION

Tropical Isles, Icy Seas, and Empire

R. Scott Winters

Willam Hodges singularly transformed how 18th-century Europeans envisioned the world. Accompanying Captain James Cook's epic second voyage—arguably one of the great scientific contributions to the Age of Reason—Hodges linked imagery to the South Pacific's lore and, supervising the engravings for the official narrative, immortalized Cook's mythos. In subsequent decades, however, the connection between artist and art was broken; over time recognition of Hodges, his work, and his contributions to our understanding of the natural world faded into obscurity.

William Hodges, 1744–1797: The Art of Exploration is the first retrospective of the artist's oeuvre. The exhibition, organized by and first shown at the National Maritime Museum in London, is at Yale University through 24 April. Most of the works have not been previously displayed in the United States; some of the paintings, such as the paired pieces *A View of Matavai Bay in the Island of Otaheite* and *A View Taken in the Bay of Otaheite Peha*, have not been exhibited together since 1776 (*1*). Furthermore, it is the first time Hodges's works from the Cook expedition and his subsequent three-year tour through India have been shown together. Augmenting the exhibition is an ambitious catalog (with a foreword by David Attenborough); however, reproductions cannot accurately convey the depth or opulence of these aristocratic works.

As the official draughtsman for Cook's circumnavigation, Hodges was to accurately portray identifiable topographical features, such as coastal profiles, to complement the planimetric drawings made by the officers. But Hodges did more. Aligned closely with the expedition's naturalists (Johann and George Forster) and astronomer (William Wales), he attempted to

The reviewer is in the Division of Oncology, Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104, USA. E-mail: winters@genome.chop.edu

capture the immediacy—indeed to report—the experience of the expedition. A pioneer in plein air, he often worked outdoors, emphasizing the atmospheric and climatic conditions that surrounded him.

A View of Cape Stephens in Cook's Straits with Waterspout is an exemplar of Hodges's sublime work. This painting, executed for the Admiralty upon the crew's return, depicts the *Resolution* rounding Cape Stephens to rendezvous with its sister ship, the *Adventure*. His portrayal is experiential. The *Resolution* precariously tosses between turbulent sea and threatening sky, narrowly escaping one of four simultaneous waterspouts it encountered on 17 May 1773.

Although Hodges worked directly from nature and closely with the expedition's scientists, he did not passively record. Instead, his work reflects a dialogue among observation, scientific theory, and artistic paradigm. Hodges often took the refracted scientific observations and reconstructed them within a classical motif. Formally trained as a landscape painter under Richard Wilson (the leading British landscape painter of the day), Hodges interpreted his observations through the lens of classicism: an ideal for art, human progress, or philosophy. *A View of the Monuments of Easter Island* depicts a depauperate landscape startlingly interrupted by colossal stone totems; storm clouds billow in the background and human remains rest in the foreground. By way of this exotic scene, he guides his viewer through a social commentary: divine effigies on the now-denuded terrain allude to the former presence of a more advanced civilization. The artistic framework here—the pictorial theme of *Et in Arcadia Ego*—recurs throughout his South Pacific works, often enfolding his anthropological observations of indigenous culture such as the Tahitian *tupapau* (spirits of the dead).

Cook's second voyage sought *terra aus-*

A View of Cape Stephens in Cook's Straits with Waterspout (1776). Although Hodges probably made several sketches of the waterspouts on or soon after the afternoon the *Resolution* encountered them off New Zealand, this oil was painted for the Lords Commissioners of the Admiralty after his return to London.

tralis incognita, a mythical, inhabited southern continent. For 117 days the *Resolution* searched the southern latitudes, crossing the Antarctic Circle three times. Thousands of miles from charted coasts, the barren, monochromatic seascape of overcast perpetual light was broken only by pack ice and towering icebergs. When the *Resolution* finally moored in Dusky Bay (New Zealand) for respite, the lush and verdant landscape was a welcome relief from the frigid temperatures and meager rations. Hodges, his paint thawed, then executed *View in Pickersgill Harbour*. Therein a fallen tree bridges the ship's deck and the luxurious, almost claustrophobic, forest. Illuminated, as if divinely, in the background, the "wilds" have been cut and tamed by the crew in order to erect their astronomical observation tent and dry their laundry.

While preparing the exhibition, the curators noticed that the paint on *View in Pickersgill Harbour* is unusually thick. X-rays of the canvas revealed a complete, ghostly underpainting. The obscured image portrays two monumental icebergs in a semi-frozen sea. The left berg, a distinctively shaped top-heavy figuration, is identical to a description by Forster in his diary. This painting is graphic confirmation of Cook's ventures into the deep southern sea and is the earliest known depiction of the Antarctic.

Note
1. These two pieces, displaying opposite sides of the island off Tahiti, present male and female (war and peace) allegories of Hodges's ethnographic interpretation of empire.

Patents on Human Genes: An Analysis of Scope and Claims

Jordan Paradise, *Lori Andrews, Timothy Holbrook

Patents by their very nature limit what others can do, offering a period of exclusive rights over the invention to the patent holder in return for public disclosure of information about the patented invention so that other inventors can build on it—for example, by making a better mousetrap out of other materials. In the case of a human gene sequence, however, the "invention" is the information. Consequently, disclosure of that information does not allow others to build on it. Gene patents, especially, limit what can be done in the realm of scientific research and medical care because there are no alternatives to a patented gene in diagnosis, treatment, and research (*1–4*). When gene patents are granted improperly and in an overly broad manner, those problems are compounded.

U.S. patent law requires that subject matter be useful (*5*), novel (*6*), and nonobvious (*7*) and fulfill four basic disclosure requirements: written description, enablement, best mode, and definiteness (*8*). When a patent is issued, the patent holder gains the right to exclude others from making, using, selling, offering to sell, or importing the invention for 20 years (*9*).

Evaluating Gene Patent Claims

To gain an understanding of whether the claims contained within issued patents covering human genetic material meet the existing statutory requirements under U.S. patent law (*10*), we undertook a multiyear project overseen by an advisory board that included two geneticists, two consumer advocates, and the head of an organization that runs a nonprofit tissue bank.

Eleven project personnel (including lawyers, licensed members of the Patent Bar, law students, and molecular biologists) identified human gene patents that represented a range of genetic diseases—from single gene to multigene disorders, from diseases where the genetic predisposition

The authors are with the Illinois Institute of Technology, Chicago-Kent College of Law, 565 W. Adams, Chicago, IL 60661, USA.

*Author for correspondence. E-mail: jparadise@kent-law.edu

has been identified to those where the causal nexuses are still being identified. We used the term "human gene patent" to include not only patents on complete human gene sequences, but patents that cover any human genetic material, such as mutations in a gene, or diagnostic methods that utilize human genetic material that would effec-

> " **...something needs to be done about the number of human gene patents being granted that arguably do not measure up to the federal patent law.**"

tively preclude the use of that material by others. We chose genetic diseases that were subject to public attention and for which problems in gene patents could potentially have an impact on research and health care. The human gene patents were not chosen with any expectation that they would have problems with their claims.

The analysis was done in a rolling fashion over the period of January 2003 to May 2004, to generate a database of at least 1000 individual claims. Initially, we examined human gene patents that dealt with the main gene or genes associated with two multigenic diseases: Alzheimer's disease and breast cancer. We then examined four single-gene diseases: ataxia telangiectasia, Canavan disease, familial dysautonomia, and hereditary hemochromatosis. We then chose three diseases that were multigenic in nature and whose genetic basis was less clear, for which a number of genes have been identified as playing a key role: asthma, obesity, and schizophrenia.

A list of human gene patents for the nine selected genetic diseases was generated by means of the U.S. Patent & Trademark Office (USPTO) database, advance search option (*11*). We used the USPTO Web site because it is the publicly accessible, federal government site that scientists and inven-

tors are expected to consult in order to investigate whether something has been patented or there is an application pending. The patents that we analyzed are not an exhaustive list of human gene patents issued by the USPTO during this period, owing to the rolling time frame and our specific search terms.

For our analysis, we examined 74 relevant patents on human genetic material, all of which contained multiple claims (such as a claim over a gene or a claim over mutations of that gene). Because the USPTO and the courts examine each claim with respect to its validity, we did so as well, assessing the 1167 claims in the 74 patents. Where the project personnel felt that a specific claim did not meet one or more of the legal requirements of patentability, it was deemed problematic. These findings are the conclusions of project personnel and are not necessarily predictive of eventual validity determinations by the courts or the USPTO.

We found that 38% of claims were problematic (see table on page 1567). Some claims had multiple problems, resulting in 677 cumulative problems within the 448 problematic claims. Of the 677 total problems identified, written description and enablement/utility problems were the most frequent (see figure page 1567). Many patents claimed far more than what the inventor actually discovered. Some applicants took advantage of the redundancy of the genetic code by, for example, claiming the sequence of a protein within a patent and then also asserting rights over all of the DNA sequences that encode for that protein without describing those DNA sequences.

Some patents exhibited written description problems by claiming discoveries the patent holder did not specifically describe. One patent covers not only the particular polymorphism the inventor discovered but all other polymorphisms discovered in the future by anyone else in a region encompassing over 12 mega–base pairs (Mbp).

Other patent claims were problematic with respect to utility. In one patent, the inventor had shown how a polymorphism could be used to predict asthma. The inventor additionally claimed various uses of the polymorphism to predict other conditions, although the inventor did not show that the polymorphism was linked to those conditions.

Another troubling issue arose where a claim was drafted disclosing only a correlation between two things, often the presence of an isoform or mutation and some multigenic disorder or a disorder having a genetic component. In some cases, the patent holder did not describe how the correlation was used to predict the disease. One such patent claims a method of detecting whether a subject is at increased risk of developing late-onset Alzheimer's disease, comprising directly or indirectly detecting the presence or absence of a particular protein isoform. The claim does not account for other factors that relate to a person's developing Alzheimer's disease and does not indicate the specific relation between the presence of the isoform and the increased risk, but merely provides that there is a correlation. In fact, even where a patient does not have the isoform, he or she could still develop the disease, and those with the isoform might never develop the disease. The "indirect" detection method is also vague enough that it could include any diagnosis based on external factors as well, such as memory loss (a diagnostic method that was not invented by the patent applicant).

We also found patent claims that suffered from one or more problems but were saved from being classified as problematic by the drafting language. For example, one claim reads, "(t)he method of claim 1, 2, or 3, wherein the method further comprises amplifying the sequence-altered PAH DNA by use of the polymerase chain reaction (PCR)." Two of the three referenced claims were problematic. Claim 1 had written description, enablement/utility, and novelty/nonobviousness problems and claim 3 had enablement/utility problems. Project personnel concluded that this claim was not problematic, however, because it referenced claim 2, which contained no problems with any of the established categories. This language may create a chilling effect on researchers who want to use methods listed in claims 1 and 3 of the patent, but do not realize that the patent is open to challenge as not validly covering those methods.

Discussion

Prior studies have found that gene patents have the potential to deter medical research and health care (*12*, *13*). A variety of policy alternatives are being considered to remedy these negative effects. These include ban-

BREAKDOWN OF PROBLEMS IDENTIFIED IN HUMAN GENE PATENTS FROM NINE GENETIC DISEASES	
Total patents	74
Total claims	1167
Total claims with problems	448
Cumulative problems	667

ning patents on genes entirely or narrowing their scope, exempting researchers and diagnosticians from the reach of the patents, creating a system of either patent pools or compulsory licensing, recognizing the rights of third parties to oppose the granting of a particular gene patent, and allowing the people who are the sources of the patented genes to have a greater say in their use (*4*).

Another avenue may be to have the USPTO remedy internal problems that lead to the issuance of gene patents that arguably do not comply with existing patent law. This is not a new problem or one that is specific to gene patents (*14*, *15*). As with any new technology, the USPTO must have competent patent examiners to guarantee that patents are not issued that are overly broad or overarching. For example, examiners in the newly created business method patent category were criticized in the late 1990s for granting poor-quality, overbroad patents lacking novelty, all of which are problems attributed to improper technical backgrounds, inadequate training, and financial incentives. Because some examiners within the patent office may not be familiar with DNA-based technologies, one potential remedy may be more training or special selection of patent examiners from a related educational background (*14*).

Some have even argued that applications should be reviewed by the USPTO with different levels of scrutiny, depending on how much social cost they entail (*14*). Reports indicate that on average, the total time spent by a patent examiner on a patent application is about 18 hours (*16*). With gene patent applications often involving extensive biological sequence information for each individual claim, it may be that adequate time is not being invested in thoroughly investigating the patentability of the claimed material. Where the enforcement of a patent has the potential to be so costly to society in terms

Analysis of problems observed within claims.

- 13.1%
- 37.5%
- 6.9%
- 42.4%

Legend:
- Written description
- Enablement/utility
- Novelty/nonobviousness
- Definiteness

of medical research, health care, and downstream innovation because there are no adequate substitutes, safeguards could be installed to ensure that the application is examined more closely (*14*).

The USPTO could also revamp financial incentives to promote decisions based on the quality of patents rather than their quantity. Currently, patent examiners are encouraged with monetary bonuses to grant patent applications, a policy that has the unsettling effect of rewarding examiners for quickly pushing patents through the patent office. Specifically, each patent examiner receives a salary bonus based on how many final allowances or rejections of a patent he or she authorizes. Because a rejection can be challenged and may not become final for quite some time, it is easier to receive a bonus by allowing patents (*14*). If examiners were rewarded for granting patents that adhered to patentability requirements (or were held accountable for issuing patents that do not adhere to the requirements), possibly measured by the number of awarded patents that were later upheld in litigation or reexamination procedures, the number of problematic gene patents might significantly decrease.

Whether through amendments to the patent law, alternative licensing mechanisms, or policy changes in the USPTO itself, something needs to be done about the number of human gene patents being granted that arguably do not measure up to the federal patent law.

References and Notes:
1. See M. K. Cho, in *Preparing for the Millennium: Laboratory Medicine in the 21st Century* (AACC Press, Orlando, FL, ed. 2, 1998), pp. 47–58.
2. J. F. Merz, A. G. Kriss, D. G. Leonard, M. K. Cho, *Nature* **415**, 577 (2002).
3. E. G. Campbell *et al.*, *JAMA* **287**, 473 (2002).
4. *See* L. B. Andrews, *Nature Rev. Genet.* **3**, 803 (2002).
5. 35 U.S. Code (USC) §101.
6. 35 USC §102.
7. 35 USC §103.
8. 35 USC §112.
9. 35 USC §154.
10. Specific information regarding the patents analyzed is available from the authors upon request.
11. http://patft.uspto.gov/netahtml/search-adv.htm
12. K. Blanton, *Boston Globe*, 24 February 2002, p. 10.
13. S. Gad *et al.*, *J. Med. Gen.* **38**, 388 (2001).
14. *See* R. P. Merges, *Berkeley Tech. Law J.* **14**, 577 (1999).
15. J. H. Barton, *Science* **287**, 1933 (2000).
16. M. A. Lemley, *Northwestern Univ. Law Rev.* **95**, 1495 (2001).
17. Supported by grant DE-FG02-02ER63460 from the Office of Biological and Environmental Research, the Office of Science, U.S. Department of Energy (DOE); and The Robert Wood Johnson Foundation Investigator Awards in Health Policy Research Program. We thank the following research assistants for their dedicated work analyzing the patents: M. Bauer, T. Ebersole, C. Mitchell, G. Prihar, M. Timm, A. Rooke, and G. M. Zinkl. We are also grateful to L. Rosenow for her role as an investigator in the DOE grant and to C. G. Janson who acted as an advisor to the project.
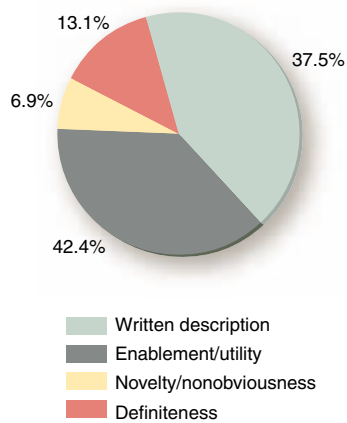
PERSPECTIVES

Information from Seismic Noise

Richard L. Weaver

It is commonly supposed that noise obscures, but does not contain, useful information. Intuition suggests that multiple scattering of waves garbles them into illegibility. Yet insights arising out of a branch of physics called "mesoscopic physics" are challenging this assumption. Theory shows that, regardless of scattering, linear waves preserve a residual coherence. This coherence leads to behaviors that confound intuition, such as Anderson localization in which a multiply scattered wave field is confined to a finite volume and unable to diffuse.

Such residual coherences can also be useful in seismology, as shown by Shapiro *et al.* on page 1615 of this issue (*1*). The authors have analyzed seismic noise to obtain new information on the structure of Earth's crust. By correlating the data from a month of ambient noise [due in part to wave-wave interactions in the ocean (*2*)] detected by 62 long-period seismograph stations in southern California, they determined the seismic response that they would have obtained from Earth's crust if they had applied forces at each of their stations. In particular, they measured the times that it took for seismic surface waves to propagate between every pair of stations. They then used tomographic techniques to create a map of seismic wave velocity with an unprecedented horizontal resolution of 75 to 100 km. The map is consistent with presumed geologic structures to a depth of 20 km. As new high-density seismograph networks come online, such results can be extended throughout the United States.

Correlation of seismic noise is a new and intriguing tool with numerous possible applications. Examples include oil exploration without explosives or thumper trucks, seismic wave profiling and deep Earth tomography from arbitrary positions without waiting for an earthquake, and the extraordinary pleasure of using and interpreting a wealth of data that were

The author is in the Department of Theoretical and Applied Mechanics, University of Illinois, Urbana, IL 61801, USA. E-mail: r-weaver@uiuc.edu

previously considered worthless.

The term "mesoscopic" is taken from low-temperature electronics, where electrons remain quantum mechanically coherent over the almost macroscopic intervals needed for electronic transport in modern small devices. Constructive and destructive interferences of the electron wave lead to a wealth of fascinating phenomena. For

Using noise in seismology. When a diffuse wave field is generated by distant sources and/or by multiple scattering, detectors report random signals. Occasionally a ray (for example, the one shown in red) passes through both detectors. As a result, the signals are weakly correlated.

example, mesoscopic fluctuations of electronic conductance affect the electronic properties of the devices. The behaviors are not confined to quantum mechanical systems, but are a consequence of linearity and of the constancy in time of the structures. Related phenomena have been observed for acoustic, seismic, and optical waves (*3*).

A transient seismic source such as an earthquake often causes two different sets of seismic waves: a main wave that propagates directly from the source, and a long-duration noisy "coda" consisting of waves (or rays) that have been scattered or reflected at least once. The variations in the intensity of the seismic coda with time have long been known to be characteristic of a region, but independent of the earthquake (*4*). Recently, Hennino *et al.* found

that, at least for a region in Mexico, the seismic coda has an additional property: Its energy is distributed in a characteristic way (equipartitioned) among the various types of seismic waves (*5*). Such partitioning is a consequence of multiple scattering. The observation thus indicates that coda waves have been scattered several times.

In the case of multiply scattered electrons and visual light, residual coherences are generally manifested in intensity correlations. At the lower frequencies of microwaves, acoustics, and seismology, we can measure fields as well as intensities. This permits observation of additional effects of residual coherence. For example, time-reversal imaging (*6*) depends on the coherence between an acoustic process and its time-reversed form, even if multi-

ply scattered. It has applications in medical ultrasound, ocean acoustics, and nondestructive evaluation of engineering structures. Another example is coda wave interferometry (*7*), which investigates changes in codas. Details of a coda waveform cannot be interpreted, but changes in a coda can correspond to changes in a medium or to the movement of scatterers within it. This method has been used to measure temperature and regularity in a body's shape (*8*), detect the growth of cracks in materials, and monitor changing environments in a volcano, mines, and a fishtank.

A third example is weak Anderson localization of seismic waves. It corresponds to an enhancement of a diffuse field's intensity at the position of its source long after the source has ceased to act. The phenom-

enon was first observed in electronics and optics, but has recently been seen with seismic waves (*9*). It is sensitive to the mean free time for a typical ray to scatter, and thus measures the elastic heterogeneity of a region.

The work by Shapiro *et al.* (*1*) arises out of helio-seismology (*10*) and ultrasonics (*11*), where it was noted that equipartitioned wave fields must have correlation functions equal to the signals that one would obtain following a concentrated impulsive force. Such correlations therefore passively reveal information about a structure that is normally obtained only by actively launching waves and detecting responses. A perfectly diffuse equipartitioned field is provided by thermal fluctuations. Correlations of thermal noise in an ultrasonic receiver circuit reveal the conventional ultrasonic waveform (*11*).

As with thermal noise, a diffuse field

generated by distant active sources also permits retrieval of the response function. It is not difficult to understand how propagation times might be revealed. A ray that is part of an isotropic diffuse field and that passes by one receiver will pass by another receiver slightly later, with its phase undisturbed except by the propagation time. Thus, the signals, although noisy, are correlated (see the figure).

Shapiro *et al.* have now demonstrated the utility of these ideas in seismology. High-resolution maps of surface wave velocity are to be expected in the near future. The prospects for other seismic applications are also good, although not yet fully proven. These and other mesoscopic phenomena may find applications in other fields of acoustics, such as ocean acoustics (*12*), room acoustics, structural acoustics and vibration, and ultrasonic nondestructive evaluation.

References and Notes
1. N. M. Shapiro, M. Campillo, L. Stehly, M. H. Ritzwoller, *Science* **307**, 1615 (2005).
2. S. Kedar, F. Webb, *Science* **307**, 682 (2005).
3. S. E. Skipetrov, B. A. van Tiggelen, Eds., *Wave Scattering in Complex Media: From Theory to Applications* (Kluwer, Dordrecht, Netherlands, 2003).
4. M. Fehler, H. Sato, *Pure Appl. Geophys.* **160**, 541 (2003).
5. R. Hennino *et al.*, *Phys. Rev. Lett.* **86**, 3447 (2001).
6. M. Fink, *Phys. Today*, 34 (March 1997).
7. R. Snieder, A. Gret, H. Douma, J. Scales, *Science* **295**, 2253 (2002).
8. O. Lobkis, R. Weaver, *Phys. Rev. Lett.* **90**, 254302 (2003)
9. E. Larose, L. Margerin, B. A. van Tiggelen, M. Campillo, *Phys. Rev. Lett.* **93**, 048501 (2004).
10. D. O. Gough, J. W. Leibacher, P. H. Scherrer, J. Toomre, *Science* **272**, 1281 (1996).
11. R. L. Weaver, O. I. Lobkis, *Phys. Rev. Lett.* **87**, 134301 (2001).
12. K. G. Sabra, P. Roux, W. A. Kuperman, *J. Acoust. Soc. Am.* **117**, 164 (2005).
13. R.L W. is supported by the National Science Foundation, grant number INT-0230312.

PLANT SCIENCES

Plant Genes on Steroids

Robert Sablowski and Nicholas P. Harberd

Plants, like animals, use steroid hormones to regulate their development through changes in the expression of target genes. However, the molecules used by plant cells to perceive and respond to the steroid signal are different from those used by animals. In animal cells, nuclear receptors generally bind to steroid hormones and directly regulate target gene expression. By contrast, in plants the steroid hormone is bound by a receptor at the cell surface. The resultant signal is then transmitted through a chain of events that include dephosphorylation of regulatory proteins and their accumulation in the nucleus. Exactly how these events are translated into changes in gene expression was, until recently, unknown. The link in the plant steroid signaling chain now is revealed by He and colleagues on page 1634 of this issue (*1*) and by Yin and co-workers in a recent issue of *Cell* (*2*).

The steroid hormone found in plants is brassinosteroid (BR). BR controls multiple processes, including cell expansion, light-induced differentiation, seed germination, and vascular development (*3*). This steroid hormone is detected by BRI1, a

leucine-rich repeat receptor kinase that spans the outer membrane of plant cells (see the figure). In response to BR, BRI1 inhibits BIN2, a protein that normally attaches phosphate groups to the nuclear proteins BES1 and BZR1. These phosphate groups tag BES1 and BZR1 for rapid destruction in the proteasome, the cellular organelle that degrades unwanted proteins. Thus, inhibition of BIN2 activity promotes the accumulation of BES1 and BZR1 in the nucleus. BES1 and BZR1

then activate selected BR-responsive genes (for example, genes encoding enzymes that relax the cell walls, thus permitting cell expansion) and repress the activity of others (such as *CPD*, which encodes CPD, a key enzyme in the BR biosynthesis pathway).

The two new studies demonstrate that BZR1 and BES1 are members of a new family of transcription factors. He *et al.* (*1*) found that BZR1 binds directly to specific sequences within the *CPD* gene and thus represses transcription (that is, the production of mRNA that subsequently directs synthesis of CPD). In addition, they identified a subset of BR-regulated genes that are probably direct targets of BZR1 and contain the BZR1 binding sequence. Yin *et al.* (*2*) showed that BES1 also binds
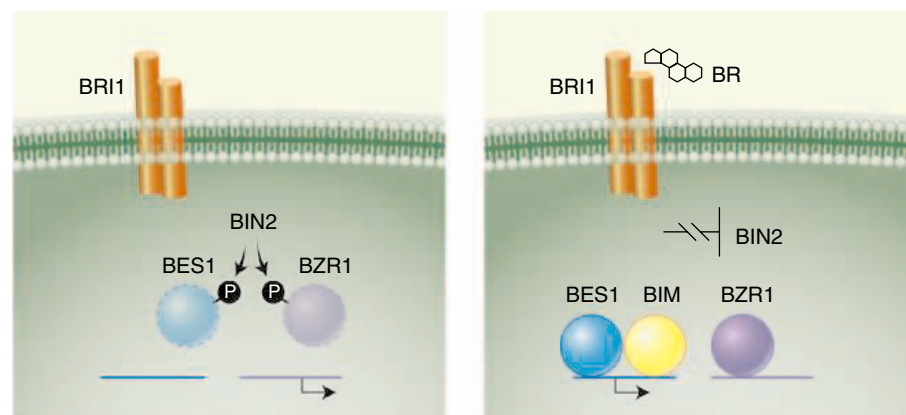
Steroid signaling in plants. (Left) In the absence of steroid, the BIN2 protein phosphorylates BES1 and BZR1, which are then degraded. Genes activated by BES1 (blue line) remain inactive, whereas genes repressed by BZR1 (purple line) are active. **(Right)** When steroid hormone is bound by the BRI1 receptor at the plant cell surface, this leads to inhibition of BIN2 and stabilization of BZR1 and BES1. BZR1 binds to target genes directly in order to turn them off, whereas BES1 acts together with BIM proteins to bind and to activate the expression of target genes.

The authors are in the Department of Cell and Developmental Biology, John Innes Centre, Norwich, NR4 7UH, UK. E-mail: robert.sablowski @bbsrc.ac.uk

directly to a target gene, but in this case acts as a transcriptional activator. BES1 does not act alone but associates with BIM proteins, which belong to a different family of DNA binding proteins. Analysis of plants carrying *bim* mutations confirmed that BIM proteins are required for a normal response to BR.

The BES1 and BZR proteins are 88% identical and, accordingly, the genetic evidence suggests that the functions of BES1 and BZR1 partially overlap (*2*). At the same time, however, they undertake specialized tasks. For example, BZR1 alone mediates feedback-regulated inhibition of BR biosynthesis (*1*). The experiments described in these papers further suggest that BZR1 and BES1 not only control different subsets of BR-responsive genes, but also have opposite effects on transcription, at least for the target genes tested. It is not clear whether these differences result from interactions with different partner proteins (it remains to be tested whether BZR1 also interacts with BIM proteins) or from recognition of different DNA sequences (the target sequence for BES1 is not as well defined as that for BZR1). Answering these questions will be important for understanding how target genes respond in a coordinated way to BR.

The repression of *CPD* by BZR1 is of particular interest: BR is thought to act in the tissues where it is synthesized, so repression of its own synthesis is probably important for controlling the range and duration of BR responses. Furthermore, such "feedback" regulation of plant hormones (phytohormones) by specific hormonal signaling pathways may be a general feature of phytohormonal biology. For example, the transcripts encoding key enzymes in gibberellin (GA) biosynthesis are regulated in a "feedback" loop by the GA signaling pathway (*4*). Another feature of the BR-response pathway that is shared with other known response pathways is that regulation is mediated by controlling the stability of key transcriptional regulators (*3*).

Together with another recent study (*5*), the work by He *et al*. and Yin *et al*. gives us an almost complete picture of how the BR signal is translated into changes in target gene expression. Although the work on BZR1 and BES1 has added the final link in the chain, the new study by Kinoshita *et al*. (*5*) reveals the first step in the process. These authors show that BR binds directly to BRI1 in vivo. Thus, the BR signaling pathway, along with the ethylene-response pathway (*6*), has become one of the best-understood hormone response pathways in plants.

Why is it so important to understand these signaling pathways? Phytohormones are central to plant development, and there are likely to be few aspects of development in which they are not involved. Understanding the developmental role of phytohormones is complicated, because each phytohormone controls so many responses. Furthermore, a single response is controlled by multiple hormones that may act either antagonistically or cooperatively. So far, the signaling pathways for the different plant hormones appear to be mostly independent of one another. This suggests that much of the integration of different signals may occur further downstream, at the level of gene regulation. Indeed, such integration has recently been shown for BR and auxin in the control of seedling growth (*7*), although a large number of genes seem to be regulated by BR or auxin independently (*8*). In addition, growth regulation by GA and auxin is, at least in part, integrated by activity of the DELLA proteins, which are putative transcriptional regulators (*9*). Identification of the key transcription factors controlled by each hormone and how they interact with each other and with their target genes will be crucial for understanding how phytohormones orchestrate plant development. The He *et al*. and Yin *et al*. studies are an important step in that direction.

References and Notes
1. J.-X. He *et al., Science* **307**, 1634 (2005); published online, 27 January 2005 (10.1126/science.1107580).
2. Y. Yin, D. Vafeados, Y. Tao, S. Yoshida, T. Asami, J. Chory, *Cell* **120**, 249 (2005).
3. Z. Y. Wang, J. X. He, *Trends Plant Sci.* **9**, 91 (2004).
4. D. E. Richards, K. E. King, T. Ait-ali, N. P. Harberd, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 67 (2001).
5. T. Kinoshita *et al., Nature* **433**, 167 (2005).
6. J. I. Alonso, A. N. Stepanova, *Science* **306**, 1513 (2004).
7. J. L. Nemhauser, T. C. Mockler, J. Chory, *PLoS Biol.* **2**, 1460 (2004).
8. H. Goda *et al., Plant Physiol.* **134**, 1555 (2004).
9. X. D. Fu, N. P. Harberd, *Nature* **421**, 740 (2003).
10. R. S. and N. P. H. are supported by the Biotechnology and Biological Sciences Research Council and the European Union.

10.1126/science.1110534

CELL BIOLOGY

Does Notch Take the Sweet Road to Success?

John B. Lowe

In the fruit fly *Drosophila melanogaster*, loss of one of the two copies of the *Notch* gene leads to the appearance of abnormal notches at the wing margins (*1*). The altered wing shape observed in such *Notch* heterozygous flies reflects the importance of Notch—and its homologs in worms, mice, and humans—for regulating key events during embryonic development (*2*, *3*). Notch family members are cell surface receptors that direct developmental processes by interacting with Notch ligands. As inferred from the phenotype of *Notch* heterozygous flies (which express only half the amount of Notch protein expressed by wild-type flies), Notch function and its developmental consequences are exquisitely sensitive to the degree of its expression. (*2*). Given that Notch is crucial for animal development, it is not surprising that elaborate mechanisms have evolved to control expression of the *Notch* gene (*4*). On page 1599 of this issue, Okajima *et al*. (*5*) unveil a new mechanism for controlling Notch expression, which depends on regulating the amount of Notch at the cell surface. These investigators report that the egress of Notch from its site of synthesis in the endoplasmic reticulum (ER) and its subsequent transit to the cell surface via the secretory pathway can be modulated by the chaperone-like activity of an enzyme called *O*-fucosyltransferase-1 (OFUT1).

In the evolutionarily conserved "core" of the Notch signaling pathway (*4*), binding of a Notch ligand to the large extracellular domain of Notch catalyzes the release of an intracellular fragment that modulates the transcription of developmentally relevant target genes. The interaction of Notch with its ligands, and the corresponding signaling events that ensue, are modulated by glycan sugars. These glycans are borne by serine and threonine amino acids within some of the many epidermal growth factor (EGF)–like repeats of the Notch extracellular domain (*6*). Serines and threonines bearing only the sugar fucose characterize the simplest of the Notch glycoforms. Addition of fucose (fucosylation) to these serines and threonines is catalyzed by OFUT1 (see the figure) (*6*).

A requirement for OFUT1 in Notch

The author is at The Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI 48109–2216, USA. E-mail: johnlowe@umich.edu

function is inferred from experiments in which loss of Notch signaling is observed when OFUT1 expression is diminished or absent. Such experiments have analyzed flies that express an *Ofut1* RNA interference (RNAi) construct (*7*) that reduces OFUT1 activity, or that carry a null allele of *Ofut1* known as *neurotic* (*nti*) (*8*), or mice with a targeted deletion of *Pofut1* (*9*). The Notch-deficient phenotypes observed have been attributed to loss of productive interactions between nonfucosylated Notch and its ligands, implying a requirement for serine- and threonine-linked fucose for this interaction to take place. This interpretation rests, in part, on the assumption that Notch surface expression remains unperturbed by loss of OFUT1. Indeed, in the experiments where RNAi was used to decrease OFUT1 expression, the subcellular distribution of Notch was presumed to be normal, and Notch was thought to move correctly through the secretory pathway to the cell surface (*7*). The conclusion that fucosylation of Notch is required for its interaction with its ligands is supported by studies in which the strength of this interaction in vitro is diminished or lost altogether when Notch is in a nonfucosylated form (*6*, *10*, *11*).

Enter Okajima *et al*. (*5*), who reexamine whether OFUT1 is required for the cell surface localization of Notch. They carefully analyzed the localization of Notch in flies carrying an OFUT1 RNAi construct that decreases OFUT1 activity. These investigators found that the predominant apical surface localization of Notch in wild-type fly cells is diminished when OFUT1 is depleted from these cells and is accompanied by a compensatory accumulation of intracellular Notch. Likewise, they observed an OFUT1 dependency in the secretion of a soluble chimeric version of Notch, thereby extending earlier observations (*10*). Three other EGF-repeat–bearing proteins retain their wild-type pattern of cell surface expression in the absence of OFUT1, implying that this enzyme is specifically required for the proper localization of Notch at the cell surface.

The first reports describing OFUT1 proposed that, like other fucosyltransferases, it was localized in the Golgi lumen. This view fits with the transmembrane transport of the fucosyltransferase substrate GDP-fucose from its site of synthesis in the cytosol to the Golgi lumen. However, Okajima *et al*. (*5*) now report that OFUT1 is localized in the ER. Furthermore, they identified a carboxyl-terminal ER retention sequence in the enzyme (the tetrapeptide HEEL), and demonstrated a requirement for this sequence in the ER localization of OFUT1. They also note that in a previously described fly mutant (*Ofut1^sh^*) associated with a severe loss-of-function Notch-deficient phenotype, OFUT1 lacks the carboxyl-terminal ER-retention sequence. Okajima and co-workers conclude that mislocalization of this catalytically active mutant enzyme accounted for the Notch-deficient phenotype.

Together with the apparent absence of GDP-fucose in the ER lumen, these observations prompted Okajima *et al*. to propose that OFUT1 may control the release of Notch from the ER, perhaps through a chaperone-like activity that is distinct from its fucosyltransferase activity. To explore this possibility, the authors created and tested a mutant OFUT1 fly protein. They disabled the enzyme's fucosylation activity by replacing an invariant arginine residue (residing within a proposed GDP-fucose binding site) with alanine. When expressed on a fly *Ofut1*-deficient background, this mutant suppressed intracellular Notch accumulation characteristic of OFUT1 deficiency, and restored Notch to the cell surface. Further confirmation that Notch fucosylation per se does not control OFUT1-dependent Notch secretion came from their observa-

- OFUT1
- R245A mutant OFUT1 (fucosyltransferase-deficient)
- Fucose
- Newly synthesized unfolded Notch
- Hypothetical Notch folding intermediate
- Improperly folded Notch
- Fully and properly folded Notch
- Fully and properly folded, *O*-fucosylated Notch
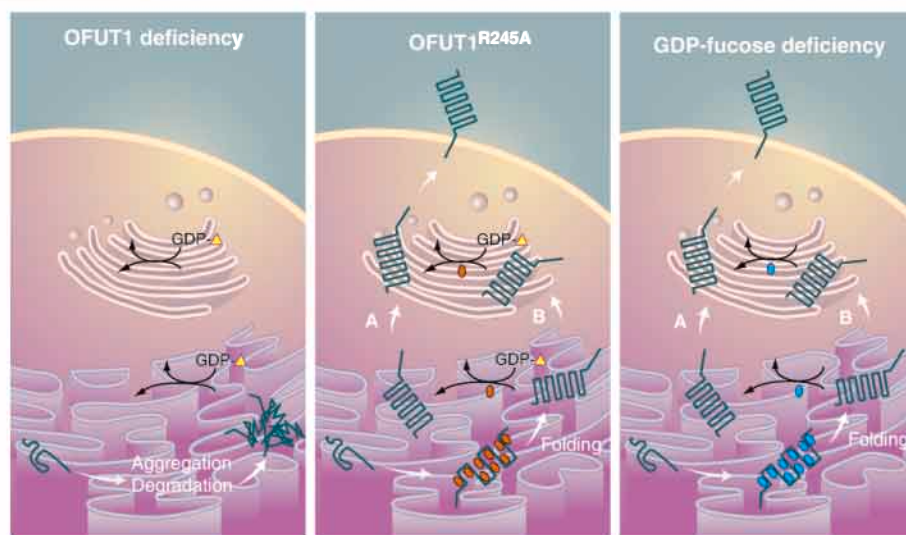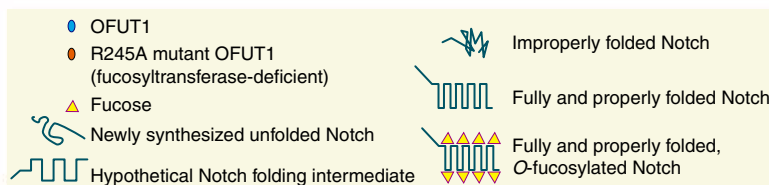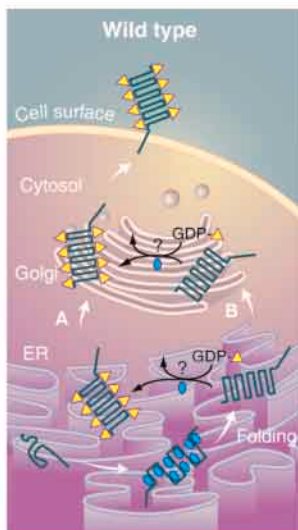
OFUT1 leads the way. (Top) In wild-type cells, newly synthesized Notch undergoes efficient folding, aided by OFUT1 (*5*). How OFUT1 contributes to Notch folding and ER egress is unknown. OFUT1 also catalyzes the fucosylation of serines and threonines on Notch in a GDP-fucose–dependent manner. It is not yet clear whether *O*-fucosylation of Notch takes place in the ER with subsequent transit of Notch to the Golgi (A), or in the Golgi following transit of nonfucosylated Notch from the ER (B), or in both compartments. Post-Golgi trafficking events move correctly folded *O*-fucosylated Notch to the cell surface, where it is able to bind to its ligands. **(Left)** In the absence of OFUT1, Notch accumulates in the ER, perhaps in aggregates composed of unfolded and misfolded Notch isomers. Cell surface expression of Notch is negligible, thereby accounting for the Notch phenotype in flies with OFUT1 deficiency. **(Middle)** In contrast, expression of a "fucosyltransferase-dead" OFUT1 mutant (the R245A missense allele) in flies effectively suppresses ER accumulation of misfolded or unfolded Notch observed in OFUT1 deficiency, and restores cell surface expression of Notch. Under these conditions, Notch is not fucosylated, revealing that OFUT1 contributes to Notch folding and ER egress in a fucosylation-independent manner. **(Right)** Likewise, in GMD deficiency, where GDP-fucose synthesis is largely abolished, OFUT1 supports normal folding of Notch, its egress from the ER, and normal cell surface expression, even though Notch is not *O*-fucosylated. Apparently, *O*-fucosylation of Notch is required for appropriate activation of Notch signaling under some conditions, but not others (*5*, *8*).

tion that Notch moved to the cell surface and did not accumulate inside fly cells deficient in GDP-mannose 4,6-dehydratase (*Gmd*), an enzyme required for the synthesis of GDP-fucose.

The inference that OFUT1 has a fucosylation-independent chaperone-like activity that promotes release of Notch from the ER is supported by several findings. For example, a fucosylation-deficient form of OFUT1 promotes the secretion and ligand-binding activity of mutant forms of Notch that are otherwise impaired. Taken together, the evidence supporting the authors' assertion that a Notch-specific chaperone activity is embodied in a glycosyltransferase is strong, and unprecedented.

These observations raise many intriguing questions. What cofactors does OFUT1 require or recruit to help it complete its task of enabling Notch to exit the ER? Does it operate as part of a protein complex? Is the activity of OFUT1 analogous to the domain specificity of the ER chaperones, Boca and Mesd, which are required for folding and transport of low-density lipoprotein (LDL)

receptor family members, or does it have broader effects? To what extent is the chaperone-like activity of OFUT1 specific for Notch? Does it also control the release of other EGF domain–containing proteins? In that regard, it may be important to revisit earlier reports of phenotypes associated with OFUT1 deficiency in flies (*8*) and mice (*9*), and to learn why nonfucosylated Notch is apparently functional in some situations but not others.

Given the observation that OFUT1 transcripts vary during *Drosophila* development (*7*), we need to determine if this variation is sufficiently dynamic to effect another layer of physiological control over Notch-dependent development. Finally, as suggested by Okajima *et al.*, it will be crucial to discern whether their observations represent the leading edge of a new paradigm in which ER-localized glycosyltransferases contribute to quality-control events. Recent studies (*12*) suggesting that *O*-mannosyltransferase (Pmt2p) in yeast is involved in protein folding should be extended to discover whether or not the mannosyltrans-

ferase activity of this enzyme accounts for its chaperone-like activity. In this regard, the ER-localized glycosyltransferases discussed by Okajima *et al.* merit further attention.

References and Notes
1. O. L. Moohr, *Genetics* **4**, 252 (1919).
2. S. Artavanis-Tsakonas, M. D. Rand, R. J. Lake, *Science* **284**, 770 (1999).
3. G. Weinmaster, C. Kintner, *Annu. Rev. Cell Dev. Biol.* **19**, 367 (2003).
4. T. Kadesch, *Curr. Opin. Genet. Dev.* **14**, 506 (2004).
5. K. Okajima, A. Xu, L. Lei, K. D. Irvine, *Science* **307**, 1599 (2005); published online 3 February 2005 (10.1126/science.1108995).
6. N. Haines, K. D. Irvine, *Nat. Rev. Mol. Cell Biol.* **4**, 786 (2003).
7. T. Okajima, K. D. Irvine, *Cell* **111**, 893 (2002).
8. T. Sasamura, *Development* **130**, 4785 (2003).
9. S. Shi, P. Stanley, *Proc Natl. Acad. Sci. U.S.A.* **100**, 5234 (2003).
10. T. Okajima, A. Xu, K. D. Irvine, *J. Biol. Chem.* **278**, 42340 (2003).
11. F. Ahimou, L. P. Mok, B. Bardot, C. Wesley, *J. Cell Biol.* **167**, 1217 (2004).
12. K. Nakatsukasa *et al.*, *J. Biol. Chem.* **279**, 49762 (2004).
13. J.B.L. is supported by the HHMI, and NIH grant number 1P01CA71932.

SIGNAL TRANSDUCTION

A New Mediator for an Old Hormone?

Sylvia Curtis Hewitt, Bonnie J. Deroo, Kenneth S. Korach

Estrogens regulate a remarkable array of physiological processes in mammals, including but not limited to reproduction, cardiovascular health, bone integrity, and behavior. Since the isolation of estrone in 1929, understanding how estrogen carries out its functions has captivated the research interests of many scientists and clinical investigators. Early postulates that estrogen acted as a metabolic cofactor were dispelled by the discovery of steroid receptor proteins (*1*). From this emerged the field of nuclear hormone and orphan receptor research. Subsequent studies by a number of investigators indicated that steroid receptor proteins, which are transcription factors localized in the nucleus, affect the expression of target genes when bound to a steroid hormone. Current research suggests, however, that this may not be the only mechanism of estrogen action. Instead, estrogen may acti-

vate diverse signaling pathways, some of which are independent of estrogen receptors (see the figure). On page 1625 of this issue, Revankar *et al.* (*2*) contribute to this body of knowledge with their investigation of a new estrogen signaling pathway in human cells.

The mechanism responsible for many diverse effects of estrogen involves the proteins to which estrogen binds, the estrogen receptors (ERs). In this "classical" or "genomic" mechanism, estrogen molecules diffuse into the cell and bind to the ER, which exists as two different forms, ERα and ERβ. The binding of estrogen to either of its receptors induces a conformational change in the ER, resulting in its interactions with the transcriptional coregulators and regulatory DNA sequences of target genes. The resulting fluctuations in mRNAs and the proteins they encode underlie an overall physiological response that takes place within hours following estrogen exposure.

Many physiological estrogen responses occur in this time frame. However, there are rapid biochemical and physiological responses that occur within seconds or min-

utes after estrogen administration that cannot be accounted for by changes in gene expression mediated by nuclear ERs. Ten years before the cloning of the nuclear ER in 1987, a second high-affinity estrogen-binding activity associated with rat uterine tissue was reported (*3*, *4*). This membrane-associated estrogen-binding activity was not the focus of many research groups, and it was often dismissed as possible contamination of membrane preparations. Some researchers did attempt to identify this binding activity, but it proved difficult to isolate, and pharmacological compounds that would definitively isolate a non-nuclear pathway were not available. Subsequent advances in experimental techniques resulted in further progress toward understanding these unexplained physiologic responses. Thus, models of estrogen action that used membrane-mediated estrogen signaling began to evolve. These models are based on data from various studies, and their diversity arises because of the research fields they represent, as described in recent reviews (*5*, *6*).

These "rapid response" models of estrogen action can be grouped into two types. First, some models propose that the rapid membrane events are mediated by the classical ERα, which initiates signaling cascades by associating with membrane structures including G proteins, caveolins, and receptor tyrosine kinases. Within the published literature describing this mechanism, there are discrepancies concerning which

The authors are at the Laboratory of Reproductive and Developmental Toxicology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA. E-mail: korach@niehs.nih.gov

molecules and pathways interact and which signals result. This mechanism is favored by investigators who study "traditional" estrogen signaling. In the second type of model, the ERα is not involved and another membrane-associated estrogen-binding protein is believed to mediate the response to estrogen. Because this mechanism has been investigated by fewer laboratories, progress has lagged behind research into ERα-mediated signaling. However, a recent report by

"Classical" and "rapid response" models of estrogen action. In the classical pathway of estrogen action, estrogen binds to the estrogen receptor (ER), a ligand-activated transcription factor in the nucleus that regulates gene transcription by binding to DNA regulatory sequences. Rapid effects of estrogen also may occur through at least two possible mechanisms: The ER located at the plasma membrane may act coordinately with other membrane proteins, leading to initiation of signaling cascades. Alternatively, non-ER membrane-associated proteins (green) that bind to estrogen, such as GPR30, may also trigger a rapid intracellular response (2, 6). Both mechanisms lead to alterations in gene transcription.

Thomas *et al.* (7) as well as findings reported here by Revankar and colleagues (2) redress this imbalance. These groups describe the properties of GPR30, a G protein–coupled receptor (GPCR) with characteristics that make it a candidate for the membrane estrogen-binding activity previously observed. Thomas and co-workers report high-affinity estradiol binding to the membranes of breast cancer cells, which lack ERα and ERβ but express GPR30. Membranes from human embryonic kidney cells, which lack all three molecules, acquire estradiol-binding activity when induced to express GPR30. GPR30 binding is selective for estradiol and the ER antago-

nists tamoxifen and ICI 182,780. In contrast, the weak estrogens, estriol and estrone, have very low affinities for GPR30. ICI 182,780 and tamoxifen activate GPR30-mediated responses, yet are antagonists of ER-mediated responses. The opposing activities of these compounds in model systems appear to be discrepant with the observed outcomes of antiestrogen biological activities and therapies. The ER antagonists may prove effective as tools to discern responses specific to ER or GPR30 signaling pathways.

As befits the proposed mechanism of a membrane estrogen receptor, immunocytochemistry studies by Thomas *et al.* indicate that GPR30 is present in the plasma membrane (7). In contrast, Revankar *et al.* (2) locate GPR30 in the endoplasmic reticulum membranes of COS7 monkey kidney cells expressing GPR30 linked to a fluorescent marker, as well as endogenous GPR30 in several other cell lines. Estrogen-initiated G protein signaling in the endoplasmic reticulum would be a unique mechanism to explain estrogen's many physiological effects. However, at this point, Revankar *et al.*'s observations are preliminary and disagree with previous indications that GPR30 is located in, and rapid estrogen signaling is initiated from, the plasma membrane. Although the process by which newly synthesized GPCRs exit the endoplasmic reticulum and move to the plasma membrane is not fully characterized, GPCR processing in the endoplasmic reticulum is considered critical for eventual signaling activity (8). Such processing ensures that aberrant proteins are degraded rather than exported, but can be a challenge in the characterization of "orphan" GPCRs (9), as may be the case in the current study (2).

Membrane-bound ERα has been identified in many species and tissues, but largely from cell culture models using transformed cell lines or cultured primary cells directly obtained from tissue. Examples include rat pituitary cells and endometrium, human breast cancer cells, mouse macrophages, and caveolae of sheep artery endothelial cells. Estrogen-induced rapid responses include production of nitric oxide in immortalized ovine endothelial cells, prolactin release from rat pituitary tumor cells, and an increase in intracellular calcium ion concentration in cul-

tured human endometrial cells and maturing oocytes (5). Studies with mice that lack ERα have demonstrated the physiological relevance of genomic estrogen signaling. Mice lacking both ERα and ERβ exhibit loss of plasma membrane ERα and ERβ from brain capillary endothelial cells. These studies show that membrane ERs are derived from the same genes as the classical ERs (10). In ERα-deficient mice, rapid estrogen responses in certain regions of the mouse brain are lost (11), whereas the rapid action of estradiol in mouse hippocampal neurons is not (12). Thus, there remain conflicting indications of the role of ERα in rapid estrogen responses.

The two new reports indicate the presence of endogenous GPR30 in immortalized human cultured cells (2, 7). However, examining the expression of GPR30 in vivo in estrogen-responsive tissues such as the uterus or mammary gland would be more indicative of a biological role for the interaction of estrogen with GPR30. Many of the original issues that have impeded the study of nongenomic estrogen signaling still remain. Contributing to the complexity is the diverse array of physiological consequences of rapid estrogen responses that have been reported and studied in very different biological models using dissimilar approaches. The reports of a new specific estrogen target, GPR30, may constitute a step forward by provoking studies that might lead to definitive determination of the physiological role of this potential non-ER estrogen response mediator and whether its early rapid role interfaces with later genomic actions mediated by nuclear ERs. Most importantly, the new work should eventually lead to elucidation of a clear biological role for the GRP30 signaling molecule in estrogen action.

References
1. E. V. Jensen, H. I. Jacobson, in *Biological Activities of Steroids in Relation to Cancer*, G. Pincus, E. P. Vollmer, Eds. (Academic Press, New York, 1960), pp. 161–174.
2. C. M. Revankar, D. F. Cimino, L. A. Sklar, J. B. Arterburn, E. R. Prossnitz, *Science* **307**, 1625 (2005); published online 10 February 2005 (10.1126/science.1106943).
3. R. J. Pietras, C. M. Szego, *Nature* **265**, 69 (1977).
4. T. C. Pappas, B. Gametchu, J. Yannarello-Brown, T. J. Collins, C. S. Watson, *Endoc.* **2**, 813 (1994)
5. A. W. Norman, M. T. Mizwicki, D. P. Norman, *Nature Rev. Drug Discov.* **3**, 27 (2004).
6. A. C. Cato, A. Nestl, S. Mink, *Sci. STKE* **2002**, re9 (2002).
7. P. Thomas, Y. Pang, E. J. Filardo, J. Dong, *Endocrinology* **146**, 624 (2005).
8. S. Terrillon, M. Bouvier, *EMBO Rep.* **5**, 30 (2004).
9. M. Lu, L. Staszewski, F. Echeverri, H. Xu, B. D. Moyer, *BMC Cell Biol.* **5**, 34 (2004).
10. M. Razandi, A. Pedram, I. Merchenthaler, G. L. Greene, E. R. Levin, *Mol. Endocrinol.* **18**, 2854 (2004).
11. I. M. Abraham, M. G. Todman, K. S. Korach, A. E. Herbison, *Endocrinology* **145**, 3055 (2004).
12. H. N. Fugger, A. Kumar, D. B. Lubahn, K. S. Korach, T. C. Foster, *Neurosci. Lett.* **309**, 207 (2001).

INTRODUCTION

Minerals over Mars

Over the past several years, Mars has been probed with a variety of spacecraft, robotic rovers, and telescopes. Three satellites are now in orbit about the planet, and two rovers are still engaged in exploring its surface. As a result, Mars is starting to reveal many of its secrets, including a more complicated history involving water than many expected (see the Viewpoint by Paige, p. 1575, and the Book Review by Hamilton, p. 1564). Fundamental to an understanding of its history is an accurate geologic map of the surface. This involves identifying specific minerals from space that compose rocks, soil, and ice, and then tying these results to outcrops and landforms, including those that are being analyzed in greater depth by the rovers. Thanks in particular to Mars Global Surveyor and Odyssey, and now Mars Express, this is happening.

Mars Express, operated by the European Space Agency, takes its name from its rapid assembly. It arrived at Mars just over 1 year ago, in December 2003, when it released the ill-fated Beagle II lander. Since then, it has been in a polar orbit about Mars, which provides full coverage of the planet. Mars Express is equipped with a variety of remote-sensing instruments, including several that complement and extend instruments on earlier spacecraft. In this issue, six papers present analyses collected over about 10 months of the chemistry and mineralogy of the martian surface —the prime objective of OMEGA, the visible and infrared spectrometer. The spectrometer covers the range in which many minerals uniquely absorb light, allowing their identification. Mars Express has now identified important minerals in large swaths of the planet, including the poles, at a resolution of about 100 m.

In the first research paper, Bibring *et al*. (p. 1576) highlight some of the key results and provide a global synthesis. OMEGA has mapped the distribution of three important silicate minerals, which delineate primary differences in the martian crust, and several sulfate and water-bearing minerals, which document later alteration involving water. Orthopyroxene, which is a calcium-poor silicate, is abundant in older martian rocks in the high southern hemisphere, but not in younger ones. Sulfates, including most likely gypsum, are common over a large area near the north pole (see the cover) and in layered rocks found by the two Mars rovers over a wide area. Several of these areas also contain hydrated sulfate minerals and abundant absorbed water. Generally, these minerals, indicative of water alteration, are also concentrated, but unevenly, in Mars' oldest rocks. Observations of the north polar cap itself show that large grains of water ice that are relatively free of dust are exposed as the cap wanes and its frost of dry ice sublimates during the summer.

These results, combined with those from other missions, are building up the geologic map of Mars, essentially mineral by mineral, that is needed to decipher its history. More minerals are on the way.

–**BROOKS HANSON**

Ancient Mars: Wet in Many Places

David A. Paige

New results from the Mars Express Orbiter mission reveal multiple deposits of minerals formed in the presence of liquid water. They reinforce the conclusion that ancient Mars was warmer and wetter than it is today, and increase the number of promising localities to search for evidence of past life.

We are in the midst of a veritable flood of new evidence for the past presence of liquid water on Mars. The latest deluge is emanating from the OMEGA (Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité) instrument aboard the European Space Agency's Mars Express Orbiter (Fig. 1) in the form of visible and near-infrared spectra of the martian surface. The results, published this week in *Science* (*1–6*), show evidence for multiple, localized ancient sedimentary deposits containing minerals that can only form in the presence of liquid water. They make it clear that ~3 billion years ago, water-rich environments such as those currently being explored by NASA's Mars Exploration Rovers (*7, 8*) existed in many locations.

Mars scientists have been searching for water-formed minerals on Mars for some time. The story behind OMEGA's new discoveries illustrates not only the ability of Mars to hide its secrets, but also the hit-or-miss nature of planetary exploration with spacecraft. Near-infrared reflectance spectroscopy had been a mainstay of planetary science and Earth remote-sensing methods for many decades, but it has taken many attempts to get a high-resolution near-infrared spectrometer in orbit around Mars. The French ISM Imaging Spectrometer for Mars aboard the partially successful 1988 Soviet Phobos 2 mission was the first imaging spectrometer to acquire a data set from another planet. However, ISM's low spatial resolution and limited spatial coverage prevented it from detecting the spectral signatures of water-formed minerals. In 1985, NASA had firm plans to equip its Mars Observer Orbiter with VIMS (Visual and Near Infrared Mapping Spectrometer). However, because of budgetary difficulties project-wide, VIMS was dropped from the payload well before launch of the mission, which ended in failure just before Mars orbit insertion in 1993.

University of California, Los Angeles, Los Angeles, CA 90095, USA. E-mail: dap@mars.ucla.edu

Subsequently, the next-generation French OMEGA spectrometer was launched on the Russian-ESA Mars '96 mission, which failed in Earth orbit in 1996. Like many of the Mars Express instruments, the currently operating OMEGA is a rebuilt copy of the lost Mars '96 version.

Although OMEGA has firmly established near-infrared reflectance spectroscopy as a productive approach for identifying water-formed minerals on Mars, it is not the only technique that has been successfully applied from orbit. NASA's Mars Global Surveyor (MGS) Thermal Emission Spectrometer (TES)

Fig. 1. The European Space Agency's Mars Express spacecraft orbits the planet Mars.

and Mars Odyssey Thermal Emission Imaging System (THEMIS) are both currently searching for these mineral signatures using longer wavelength thermal emission spectroscopy, a laboratory and remote-sensing technique that is also routinely applied on Earth. Their most significant discovery to date has been the identification by TES of water-formed crystalline hematite at Meridiani Terra (*9*), which was the driving motivation for the selection of the MER Opportunity Rover landing site (*8*).

With this one notable exception, the positive identification of other water-formed minerals on Mars with TES and THEMIS data sets has proven difficult, apparently because of the nature of the mineral deposits themselves. Thermal emission spectroscopy generally works best on smooth rocks containing coarse-grained minerals that produce strong thermal emission signatures (*10*). At its landing site at Meridiani Terra, the Opportunity Rover found the surface to be littered with smooth, millimeter-sized spherules of coarse crystalline hematite dubbed "blueberries," which formed from the chemical precipitation of hematite from iron-rich fluids within the sediments (*8*). The MER images clearly show that the hematite spherules are concentrated at the surface by the erosion of the fine-grained sedimentary units in which they formed (*8*), making them ideal targets for detection by TES from orbit. OMEGA has observed exposed sedimentary units associated with those explored at the Opportunity site and found spectral evidence for the hydrated sulfate mineral kieserite (*5*). This new observation is entirely consistent with the conclusion that fine-grained sediments at the MER Opportunity site contain abundant sulfate salts (*8*). Taken together, OMEGA's new results suggest the widespread exposure of water-formed minerals in fine-grained sedimentary deposits that produce thermal emission signatures that may be too weak to be detected from orbit by TES and THEMIS.

In addition to reinforcing recent conclusions regarding the presence of surface liquid water environments early in Mars's history, OMEGA is providing valuable new information about the global distribution of exposed water-formed minerals. Thus far, having mapped only 50% of the planet, and only 3% at highest resolution, OMEGA has identified hydrated sulfates and hydrated silicates almost exclusively in locations where relatively recent processes have exposed ancient materials of sedimentary origin. High-resolution MGS Mars Orbiter Camera images have revealed the presence of a much larger number of layered outcrops in a range of ancient geologic settings (*11*). The OMEGA results increase the likelihood that these outcrops contain sedimentary rocks. At martian mid-latitudes, the Mars Odyssey Gam-

ma Ray Spectrometer (GRS) has discovered extensive regions of enhanced near-surface hydrogen abundance that are estimated to contain up to 5 to 10% water by mass (*12*). OMEGA's positive detection of hydrated minerals in selected exposed outcrops increases the likelihood that some, if not most, of the mid-latitude hydrogen detected by the GRS is also in the form of hydrated minerals (*1*). The ensemble of water-bearing mineral species identified thus far on Mars, and the geological settings in which they are found, appear to require surface climatic conditions that were distinctly warmer and wetter than they are today. Determining whether these more clement environments were local or global, transient or long-lived, will require a more complete knowledge of the range of mineral species present, and their geological context.

OMEGA's discovery of multiple ancient water-rich Earth-like surface environments on Mars should redouble efforts to explore the martian surface widely for signs of past life. Later this year, NASA's Mars Reconnaissance Orbiter will be launched with a powerful suite of new remote-sensing instruments that include CRISM, the Compact Reconnaissance Imaging Spectrometer for Mars, which has more than 10 times the spatial resolution of OMEGA. If all goes as CRISM will join OMEGA to further define the extent and nature of martian water-formed mineral deposits. The future for orbital remote sensing of Mars continues to look rosy. However, as we look ahead to NASA and ESA's plans for the next decade, there are signs that a large and obvious "rover gap" is developing, given that the two or three rovers currently planned will not be sufficient to explore the diversity of scientifically exciting surface sites identified from orbit. If the successful ongoing Mars missions are teaching us anything, it is that the surface of Mars is even more diverse than had previously been thought, especially at sub-kilometer spatial scales. While the fine-scale diversity of the martian surface increases the chances of finding evidence for previously habitable or inhabited environments, it also increases the number of places that need to be searched. As best we can now determine, Mars has exposed on its surface today a record of its early environment that is better preserved than, and at least as interesting as, that of Earth. If true, then achieving equivalent levels of insight into the early histories of both planets will ultimately require equivalent levels of exploration and investigation.

References

1. J.-P. Bibring *et al.*, *Science* **307**, 1576 (2005); published online 17 February 2005 (10.1126/science.1108806).
2. Y. Langevin *et al.*, *Science* **307**, 1581 (2005); published online 17 February 2005 (10.1126/science.1109438).
3. Y. Langevin, F. Poulet, J.-P. Bibring, B. Gondet, *Science* **307**, 1584 (2005); published online 17 February 2005 (10.1126/science.1109091).
4. A. Gendrin *et al.*, *Science* **307**, 1587 (2005); published online 17 February 2005 (10.1126/science.1109087).
5. R. E. Arvidson *et al.*, *Science* **307**, 1591 (2005); published online 17 February 2005 (10.1126/science.1109509).
6. J. F. Mustard *et al.*, *Science* **307**, 1594 (2005); published online 17 February 2005 (10.1126/science.1109098).
7. S. W. Squyres *et al.*, *Science* **305**, 794 (2004).
8. S. W. Squyres *et al.*, *Science* **306**, 1698 (2004).
9. P. R. Christensen, R. V. Morris, M. D. Lane, J. L. Bandfield, M. C. Malin, *J. Geophys. Res.* **106**, 23873 (2001).
10. L. E. Kirkland, K. C. Herr, J. W. Salisbury, *Appl. Opt.* **40**, 4852 (2001).
11. M. C. Malin, K. S. Edgett, *Science* **290**, 1927 (2000).
12. W. C. Feldman *et al.*, *Science* **297**, 75 (2002); published online 30 May 2002 (10.1126/science.1073541).

10.1126/science.1110530

RESEARCH ARTICLE

Mars Surface Diversity as Revealed by the OMEGA/Mars Express Observations

Jean-Pierre Bibring,[1]* Yves Langevin,[1] Aline Gendrin,[1] Brigitte Gondet,[1] François Poulet,[1] Michel Berthé,[1] Alain Soufflot,[1] Ray Arvidson,[2] Nicolas Mangold,[3] John Mustard,[4] P. Drossart,[5] and the OMEGA team[6]

The Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité (OMEGA) investigation, on board the European Space Agency Mars Express mission, is mapping the surface composition of Mars at a 0.3- to 5-kilometer resolution by means of visible–near-infrared hyperspectral reflectance imagery. The data acquired during the first 9 months of the mission already reveal a diverse and complex surface mineralogy, offering key insights into the evolution of Mars. OMEGA has identified and mapped mafic iron-bearing silicates of both the northern and southern crust, localized concentrations of hydrated phyllosilicates and sulfates but no carbonates, and ices and frosts with a water-ice composition of the north polar perennial cap, as for the south cap, covered by a thin carbon dioxide–ice veneer.

Many questions about the evolution of Mars, including the nature of surface-atmosphere interactions, the water and CO_2 cycles on short to long time scales, crustal formation and evolution, and the role of water in creating and sustaining habitable environments, require understanding its mineralogy and composition. Analyses of data from OMEGA (*1*), the hyperspectral visible and infrared imaging spectrometer on board the Mars Express European Space Agency (ESA) mission (*2*), reveal a diverse and complex surface mineralogy pertinent to these outstanding Mars problems. Here, we describe key surface materials in the form of mafic iron-bearing silicates, ices and frosts, and hydrated minerals and sediments, which have been identified by OMEGA.

During the first 9 months of operations, OMEGA has acquired the spectra of more than 20 million pixels along ~100 tracks. With an

[1]Institut d'Astrophysique Spatiale (IAS), Bâtiment 121, 91405 Orsay Campus, France. [2]Earth and Planetary Sciences, Washington University, St. Louis, MO 63130, USA. [3]Interactions et Dynamique des Environnements de Surface (IDES), Bâtiment 509, 91405 Orsay Campus, France. [4]Geological Sciences, Brown University, Providence, RI 02912, USA. [5]Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique (LESIA), Observatoire de Paris, 92195 Meudon, France. [6]The OMEGA co-investigator team: Michel Berthé,[1] Jean-Pierre Bibring,[1] Stéphane Erard,[1] Olivier Forni,[1] Aline Gendrin,[1] Brigitte Gondet,[1] François Poulet,[1] Alain Soufflot,[1] Michel Combes,[2] Pierre Drossart,[2] Thérèse Encrenaz,[2] Thierry Fouchet,[2] Riccardo Merchiorri,[2] GianCarlo Belluci,[3] Francesca Altieri,[3] Vittorio Formisano,[3] Guillaume Bonello,[4] Fabricio Capaccioni,[4] Pricilla Cerroni,[4] Angioletta Coradini,[4] Sergio Fonti,[5] Volodia Kottsov,[6] Nikolai Ignatiev,[6] Vassili Moroz,[6] Dimitri Titov,[6] Ludmilla Zasova,[6] Nicolas Mangold,[7] Patrick Pinet,[8] Sylvain Douté,[9] Bernard Schmitt,[9] Christophe Sotin,[10] Ernst Hauber,[11] Harald Hoffmann,[11] Ralf Jaumann,[11] Uwe Keller,[12] Ray Arvidson,[13] John Mustard,[14] Tom Duxbury,[15] François Forget.[16] [Affiliations: [1]IAS, Bâtiment 121, 91405 Orsay Campus, France. [2]LESIA, Observatoire de Paris, 92195 Meudon, France. [3]Istituto di Fisica dello Spazio Interplanetario–Istituto Nazionale di AstroFisica (INAF), 00133 Rome, Italy. [4]Istituto di Astrofisica Spaziale–INAF, 00133 Rome, Italy. [5]University of Lecce, 73100 Lecce, Italy. [6]IKI, 117997 Moscow, Russia. [7]IDES, Bâtiment 509, 91405 Orsay Campus, France.

[8]Observatoire Midi-Pyrénées, 31400 Toulouse, France. [9]Laboratoire de Planétologie, 38041 Grenoble, France. [10]Planétologie, Université de Nantes, 44322 Nantes, France. [11]Deutschen Zentrum für Luft-and Raumfahrt (DLR), 12489 Berlin, Germany. [12]Max Planck Institut für Sonnensystemforschung (MPS), 37191 Lindau, Germany. [13]Earth and Planetary Sciences, Washington University, St. Louis, MO 63130, USA. [14]Geological Sciences, Brown University, Providence, RI 02912, USA. [15]Jet Propulsion Laboratory, Pasadena, CA, 91109 USA. [16]Laboratoire de Météorologie Dynamique, Université Paris 6, 75252 Paris, France.]

*To whom correspondence should be addressed. E-mail: bibring@ias.u-psud.fr

instantaneous field of view (IFOV) of 1.2 mrad (4.1 arc min), the pixel size depends on the altitude of observation and varies from 300 m, when imaged close to the orbit periapsis, up to 4.8 km from an altitude of 4000 km, whereas the cross-track FOV varies from 16 to 128 IFOV (1.1° to 8.8°). For each resolved pixel, OMEGA acquires the spectrum in 352 contiguous channels covering 0.35 to 5.1 μm. The spectral sampling is 7 nm, 14 nm, and 20 nm from 0.35 to 1.0 μm, 1.0 to 2.5 μm, and 2.5 to 5.1 μm, respectively. The spectral range and resolution have been chosen to allow the identification of major surface and atmospheric species by their diagnostic spectral absorption features, essentially in the solar reflected spectrum and, beyond 4 μm, in the planetary thermal emission.

The Mars Express orbit is almost polar (with an inclination of 86.6°) and highly elliptical, with a period of 6.72 hours (7.58 hours during the first 100 days). Because of the large value of the J2 term of the Mars gravitational field, the periapsis precesses in latitude by ~20° per month. The period has been selected so as to obtain a drift in latitude that enables the spacecraft in principle to cover the entire planet and to observe all regions at high resolution during the 1-martian-year nominal mission, with longitudinal (cross-track) nadir offsets. Figure 1 illustrates the evolution of the periapsis latitude as a function of time together with the illumination conditions. It indicates that, in the OMEGA data set acquired so far, the southern polar cap, observed at the very beginning of the mission, and northern latitude areas have been observed at medium resolution, with only the southern midlatitude zones covered at high resolution. The coverage of northern latitude areas will be performed from October 2004 until March 2005.

Mafic and ultramafic minerals. OMEGA can identify pyroxene and olivine; it discriminates between the high-calcium pyroxenes (HCPs, e.g., clinopyroxenes) and low-calcium pyroxenes (LCPs, e.g., orthopyroxenes) (3). HCPs dominate in low-albedo volcanic regions of Hesperian age (4), dark sand, and crater ejecta. The HCP-rich zones match (Fig. 2) the Thermal Emission Spectrometer (TES) data unit identified at medium resolution (> 3 km) as basalt rich (5). With its higher resolution capability, OMEGA identifies HCPs in dark sand and crater ejecta. TES has not detected LCP-dominated surfaces (6), except for a few outcrops (7). OMEGA has indeed identified a large number of LCP-rich areas where the high concentrations of LCP occur as moderate to bright outcrops in older, Noachian terrains. This is consistent with previous visible-infrared (vis-IR) analyses (8).

The northern plains of Mars, as well as high-latitude regions in the south, exhibit a distinct emissivity in TES data that has been interpreted as andesitic basalt requiring high-silica glass (5) or altered basalt requiring sheet silicates (9). In the OMEGA data of these regions, absorption features due to primary iron-bearing, well-ordered, crystalline igneous constituents are weak to absent. This could indicate that the surface material of these areas is dominated by mafic (or impact) glasses (10, 11) and/or was altered by a major weathering process. However, the resulting minerals, although highly oxidized, do not show enrichment in hydrated phases (see below).

Olivine was found in a few specific areas (3, 7, 12), mainly in crater floors, in crater rims in the southern highlands and in association with volcanic rocks in and around Syrtis Major. A typical olivine-rich outcrop in the southern hemisphere has a composition, derived through modeling (13), in which olivine represents ~30% by volume. The area exhibiting the highest olivine content is in Nili Fossae, as previously identified by Mars Global Surveyor (MGS)/TES (6).

Mafic-rich zones, with the presence of large concentration of olivine, were also found within the craters >20 km in diameter and in some of

Fig. 1. Evolution of the latitude of the orbit periapsis as a function of time, expressed in date, orbit number, and solar longitude, L_s, along the nominal Mars Express mission. The colors illustrate the illumination conditions at pericenter in solar elevation.
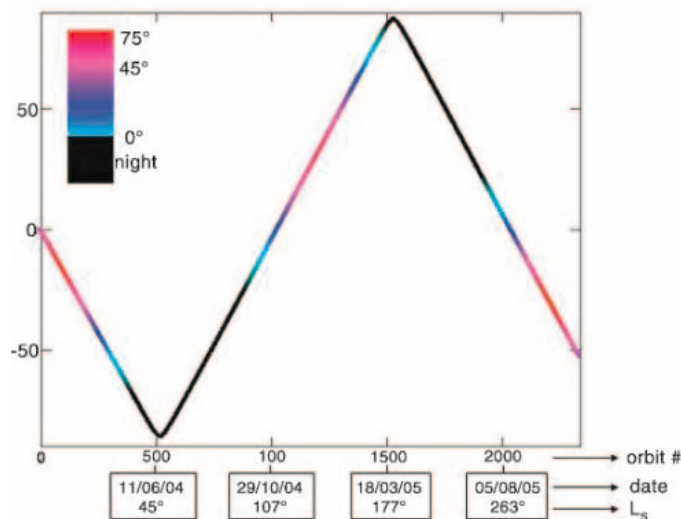
Fig. 2. Distribution of the HCPs (**left**) and LCPs (**right**) from the TES/MGS (**top**) and OMEGA/Mars Express (**bottom**) spectral imagers. OMEGA confirms the TES mapping of HCPs, obtained at moderate (>3 km) resolution, identifies and maps LCPs, and locates at higher resolution (<1 km) a variety of areas with strong mafic features.
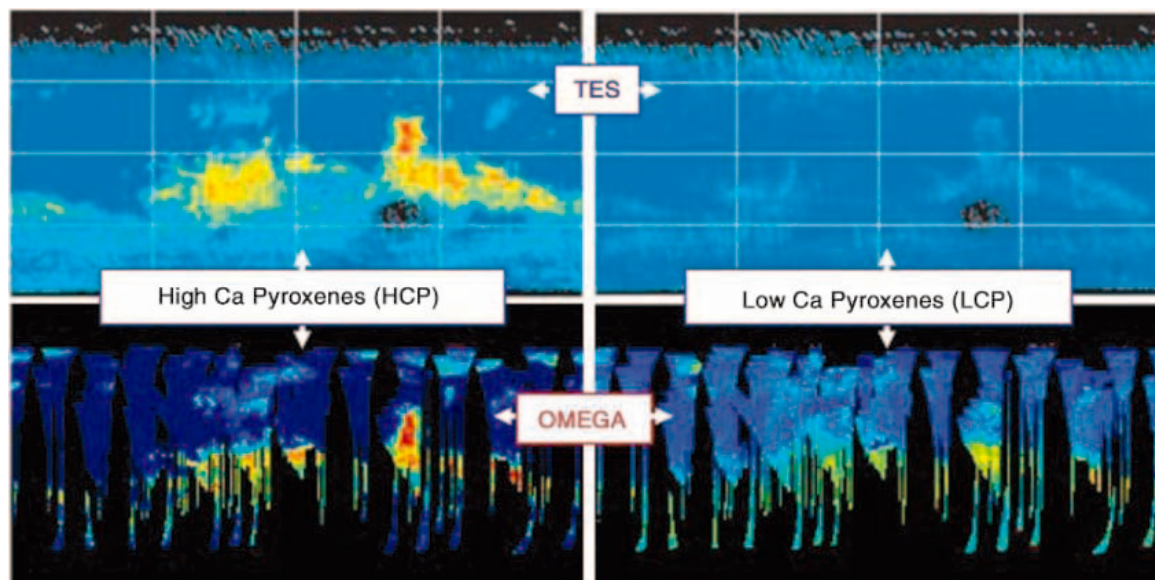
their lobate ejecta in the northern plains of Vastitas Borealis (Fig. 3). It cannot be entirely ruled out that these zones, which appear as localized aeolian deposits in the available Mars Orbiter Camera (MOC) images, have been selectively trapped inside the craters after having been transported from remote areas.
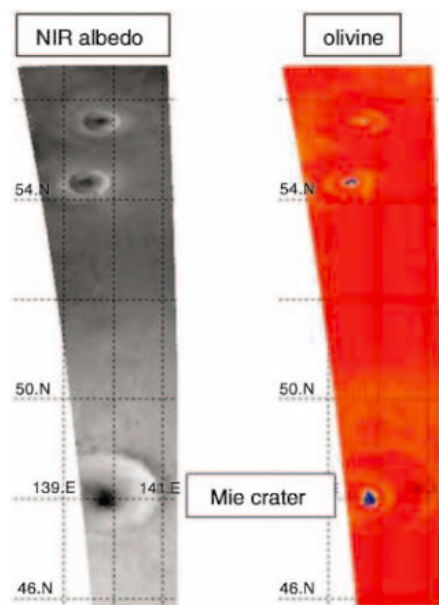
Fig. 3. Identification of olivine-rich craters within Vastitas Borealis: OMEGA coverage of an area including the Mie crater. The NIR albedo (at 1.3 μm) shows (**left**) the presence of three craters, 20 km to 50 km in diameter. The olivine map (**right**) exhibits an olivine-rich spot (in blue) within the larger crater, fainter in the smaller ones.

However, given the distances from the southern crust, a much denser atmosphere would be required. If accumulated in the early history of Mars, one would not expect these deposits to be present in most impact craters of the northern plains and not further altered, as was the surrounding material. Therefore, we consider that they constitute in-place material, excavated from beneath the northern plains cover, exposing underlying crustal and/or lava materials recently enough not to have been further altered entirely. Their composition thus reflects that of the northern crust, with an olivine abundance of up to ~30% volume.

Surface ices and frosts. During the first weeks after orbit insertion, corresponding to the late summer of the martian southern hemisphere, OMEGA identified the components of the perennial south polar cap (*14*): a thin bright CO_2 veneer, no more than some 10 m in thickness as derived from the MOC/MGS high-resolution images (*15*), covers a massive and extended glacier made of H_2O ice mixed with dust, which could not be distinguished from the surrounding materials in the previous Viking and MOC visible images. During the northern spring up to the summer solstice (20 September 2004), OMEGA monitored the defrosting of the northern areas. During this period (Fig. 4), the CO_2 frost sublimated, whereas the H_2O ice retreated, with the receding edge of the seasonal cap constituted of H_2O frost (with no residual CO_2). The CO_2 defrost ended before $L_s \sim 90°$, exposing a large water-rich residual cap (Fig. 5, right), corresponding to the bright areas mapped by the Viking Orbiter (Fig. 5, left).

By comparison with laboratory measurements of grain size and temperature effects on the near-infrared (NIR) spectra and modifying the depth and slopes of the 1.23-, 1.5-, 1.65-, and 2.0-μm absorption bands, we can assess from the OMEGA spectra the variation in surface water ice properties in terms of mean size and temperature (*16*). The water ice covering the southern cap is colder ($T < 180$ K) than that constituting the northern cap ($T > 200$ K), consistent with the TES/MGS findings (*17*). The seasonal water frost of the north cap is much finer (<100 μm) than the underlying permanent water ice (~1 mm) (*16*).

Hydrated minerals and ferric oxides. OMEGA detects hydrated minerals by the overtones and combinations of the v_1 (symmetric O–H stretch), v_2 (H–O–H bend), and v_3 (asymmetric O–H stretch) absorption features at ~1.4 μm (shallow), ~1.9 μm (deeper), and ~3 μm (large and saturated). Every OMEGA spectrum exhibits a strong 3-μm absorption, and there is no apparent systematic correlation with other compositional or morphological parameters. This is caused by the presence of a few percent of water strongly bound in minerals in the soils and surface materials (*18*, *19*). The band strengths are strongly correlated to albedo. When they are corrected for albedo (*20*), a trend appears, with the brightest soil (NIR albedo > 0.30) exhibiting the strongest 3.0-μm band (Fig. 6, A and B).

The 1.9-μm feature is found in a variety of hydrated minerals, hydroxides, phyllosilicates, and salts and serves as an indicator of these important minerals. Their precise composition is determined by diagnostic companion absorption features, such as the sharp metal-OH
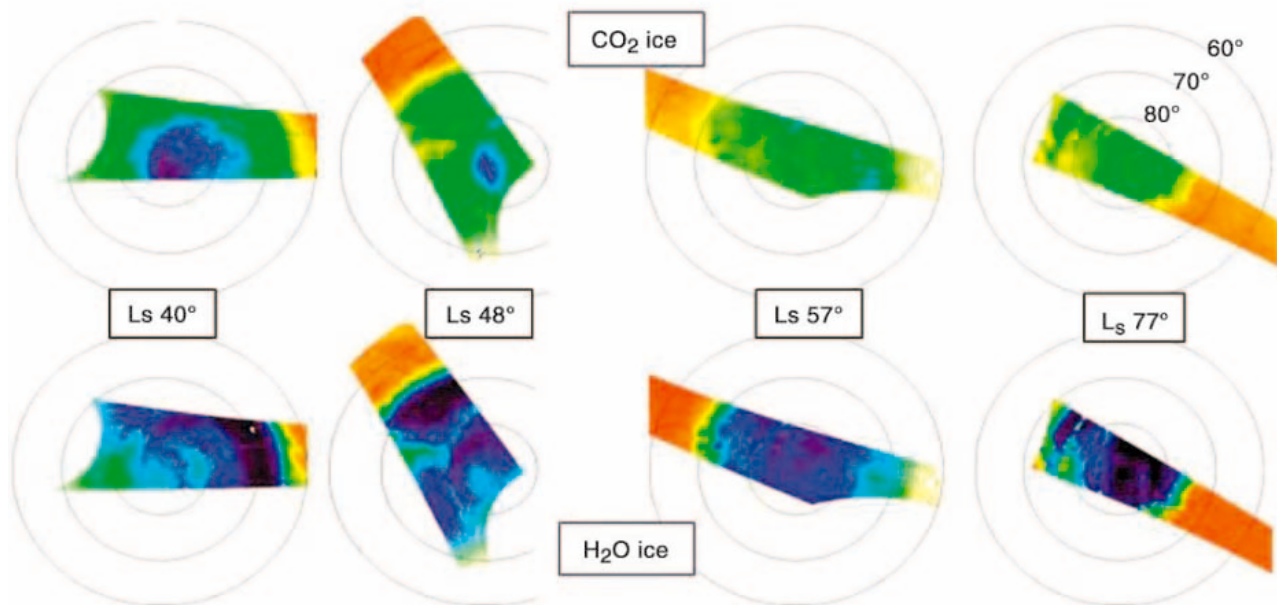
Fig. 4. CO_2 and H_2O frost evolution on the northern polar cap over the spring. These false color maps illustrate the evolution with time of the frost coverage during the northern spring, from red (0%) to dark blue (100%): The CO_2 frost regresses to leave a H_2O perennial polar cap, mapped in Fig. 5. A contribution of H_2O ice within the CO_2 ice absorption band used to build the upper images appears as the orange to green areas in the CO_2 maps.
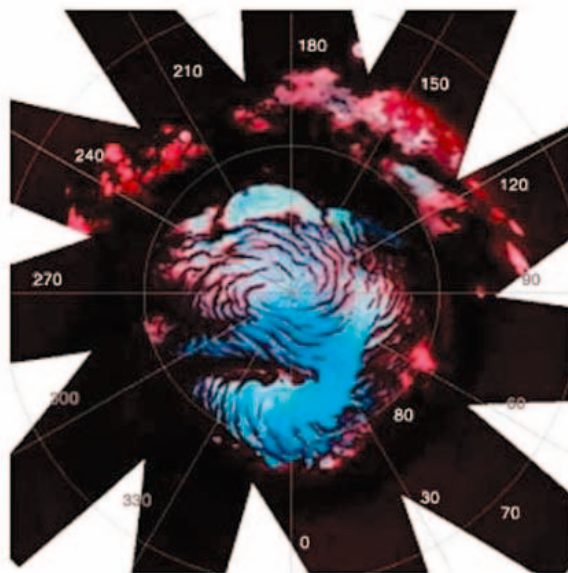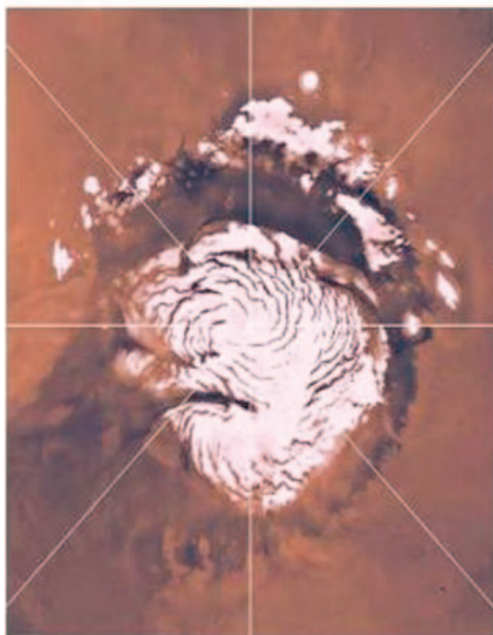
Fig. 5. Perennial northern polar cap. Viking visible image (**left**) (acquired from L_s from 110° to 155°) and OMEGA H_2O map (**right**) (acquired from L_s from 93° to 98°). The bright areas are dominated by water ice, when the CO_2 has entirely sublimated away. The blue to red false colors correspond to different water ice spectra: the blue ice is finer and colder, whereas the red one is coarser and warmer.

bands in the 2.15- to 2.4-μm region. As an example, OMEGA has unambiguously identified the mineral nontronite by its distinct 2.28-μm absorption. Nontronite, known to originate from the alteration of mafic to ultramafic igneous rocks, is the most abundant phyllosilicate detected by OMEGA (Fig. 7).

In contrast to the 3-μm feature present in all surface materials, the 1.9-μm feature tracing hydrated minerals is identified in a few specific areas only and in particular within unburied cratered units, such as those expanding out of Nili Fossae in the northern part of Syrtis Major (Fig. 6C). Enrichment in hydrated minerals is identified around an old impact crater (Fig. 6, D and E); this contrasts with the lack of detection of hydrated minerals in the volcanic outflows from Nili Patera, which buried most of the ancient terrains. These OMEGA observations point toward a hydrolytic alteration process more prevalent in the early (Noachian) history of Mars.

No enrichment in hydrated minerals is observed in lobate craters (as expected from impacts into permafrost-rich sublayers) nor within MOC/MGS-identified gully-rich slopes. Similarly, there is no definite identification of phyllosilicates within the soil covering extended areas, such as the large northern plains: This indicates that the alteration process that transformed the volcanic minerals into highly oxidized ones, such as nanophase hematite, does not seem to have led to water-rich compounds.

Sedimentary minerals. We have looked for carbonates and so far have not found areas with amounts above the OMEGA detection limit of a few percent in mass. A variety of units have been mapped: areas exhibiting hydrated minerals, layered deposits, fluvial floors, and ejecta of deep craters within Vastitas Borealis. The surface resolution of most observations is still only in the 1- to 3-km range. With CO_2 ice in the
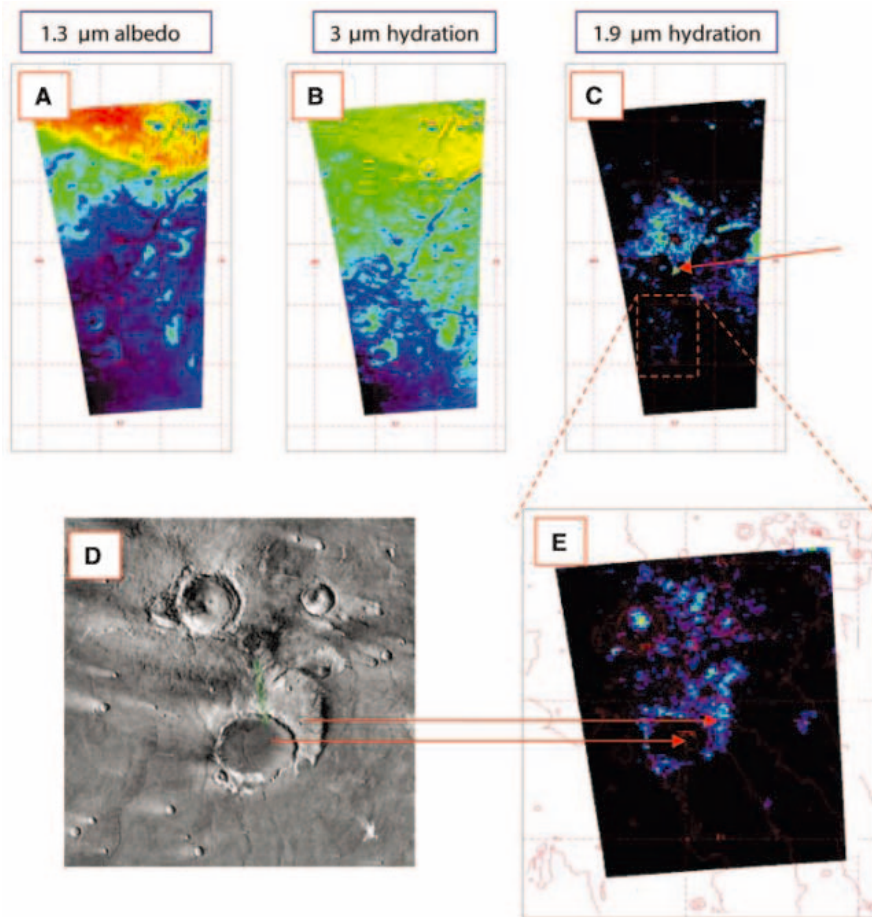
Fig. 6. Hydrated minerals; example of the Syrtis Major area. OMEGA identifies hydration (bound and adsorbed water) by the 3-μm saturated feature, in (**B**) and corrected for the lambertian albedo (**A**), and the hydrated minerals by their 1.9-μm band (**C**). The 3-μm band depth increases with albedo, whereas hydrated minerals are predominantly located in older cratered areas, as shown here in Nili Fossae. In Syrtis Major (**D** and **E**), the ancient crater ejecta exhibit hydrated minerals (red arrows), whereas the more recent lava flows are essentially free of them. The red arrow in (C) indicates the location of nontronite (Fig. 7). Color scales from blue to red indicate the range 0.05 to 0.4 in (A), 0.3 to 0.6 in (B), 0.02 to 0.06 in (C), and 0.02 to 0.05 in (E).
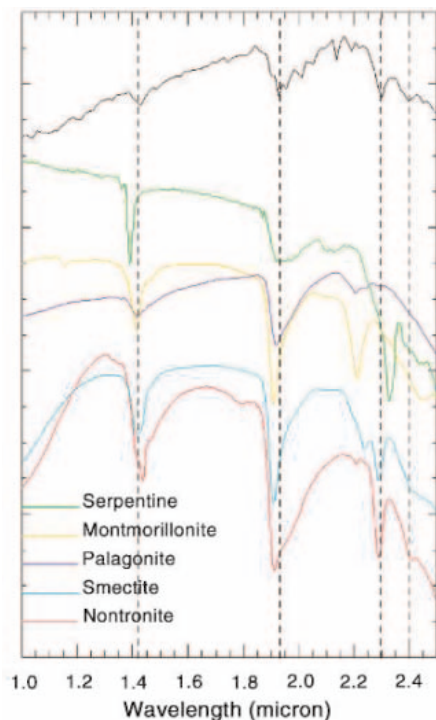
Fig. 7. Nontronite detection. The OMEGA spectrum of the area identified in Fig. 6C (red arrow) is shown in black (top spectrum), ratioed to a reference spectrum of the same session. Laboratory measurements provided by Brown University of a variety of phyllosilicates are shown below, with their diagnostic features in this spectral range. Nontronite is discriminated by its 1.42-, 1.93-, 2.29-, and 2.4-μm absorption bands.

Fig. 8. Sulfate-rich layer deposit. Within Valles Marineris, Juventae Chasma is a 2.5-km-high and ~40-km-long layered deposit, imaged in this spectacular three-dimensional view (*30*) by the HRSC/Mars Express camera. Superimposed on the corresponding black and white MOC/MGS image is the mapping of the sulfates obtained by OMEGA, with gypsum (blue) dominating in the layered strata and kieserite (red) expanding around.

perennial southern polar cap constituting a thin veneer amounting for a small fraction of the present atmospheric CO_2, the absence of carbonates, if confirmed, would indicate that no major surface sink of CO_2 is present. The initial CO_2, if indeed it represented a much higher content, would then have been lost from Mars rather than stored in surface reservoirs after having been dissolved in long-standing bodies of water. This scenario would be consistent with a view (*21*) derived from the N_2/CO_2 ratio, which is essentially identical in Mars and in Venus, and similar to the terrestrial ratio if one considers the carbonate-rich surface rocks as the sink of the initial terrestrial CO_2. If the three terrestrial bodies are assumed to have accreted a similar volatile inventory, one would not expect that at Mars two entirely different processes with CO_2 transformed into carbonates, whereas N_2, hardly precipitating as a mineral, lost by atmospheric escape, would lead to a huge depletion (> 99%) of these two constituents with almost exactly the same efficiency. The present low martian atmospheric pressure would thus originate from atmospheric loss rather than a surface sink.

OMEGA has identified hydrated sulfates through their distinct vibrational overtones involving water and hydroxyl in the 1- to 2.5-μm region. Sulfates would thus constitute the best record of a H_2O-driven past activity. Up to now, these salts have been identified by OMEGA in a variety of specific locations: within Valles Marineris and in Aram Chaos, essentially in association with light-toned layered deposits (*22*), and as large regional units in Terra Meridiani (*23*) and close to the perennial northern polar cap (*24*). Not all light-toned layered deposits, however, show evidence for sulfates, such as the White Rock area within Pollack crater: OMEGA confirms, with sub-km-resolution mapping, the lack of evidence for an aqueous origin (*25*).

Magnesium sulfates (kieserite, $MgSO_4 \cdot H_2O$) and calcium sulfates (gypsum, $CaSO_4 \cdot 2 H_2O$) have been identified. In at least one location, Juventae Chasma, kieserite appears as covering and surrounding a massive stratified deposit likely made of layers of gypsum ($CaSO_4 \cdot 2 H_2O$) (Fig. 8). In Terra Meridiani, a surface layer of sulfates is also found on etched terrains (*26*) northeast of the Opportunity landing site (OLS). It is likely that the sulfate-rich outcrops found by Opportunity in OLS belong to the same unit and were brought as patches to surface by erosion. Hydrogen enrichment was found by Gamma Ray Spectrometer/Odyssey in low latitude areas and attributed to hydrated minerals rather than subsurface ice (*27*). The hydrated minerals detected by OMEGA could be present within a much more extended subsurface layer, appearing at the surface in only a few places. This could also be the case of the wide sulfate-rich area associated with longitudinal dunes within the north polar cap (*24*): It could indicate that hydrogen-rich areas mapped by GRS at high latitude do include water ice–rich terrains and hydrated minerals spread over much wider areas, with a fraction only exposed to the surface by erosion.

As detected by OMEGA, sulfate deposits are commonly associated with hydrated silicates and/or ferric oxides at a regional and local scale; however, they are not strictly correlated at a pixel level. This would indicate that although sulfates and oxidized minerals bear a large commonality in origin, they have experienced a likely physical sorting process later.

The existence of sulfates requires water at the time of their formation. However, it does not necessarily imply that evaporation in long-standing bodies of water did occur. Other processes might have taken place, such as salt precipitation in supersaturated seeping water either evaporating or sublimating away out of water-rich ice, if the atmosphere was sufficiently tenuous, cold, and dry for water not to be stable in liquid state at the surface of the planet. Water could have been brought to the surface either as liquid or frost and made acidic by volcanic outgassing, with H_2S reacting with atmospheric O to form H_2SO_4. Sulfates would thus originate from the acidic rain or frost alteration of basalts, with olivine altered into magnesium-rich to iron-rich sulfates and HCP into calcium-rich sulfates (*28*).

Conclusions. The analyses of the NIR hyperspectral reflectance data OMEGA has acquired so far indicate that the surface of Mars exhibits a large diversity in composition, in particular at small spatial scales (~1 km). At seasonal time scales, compositional variation is dominated by the CO_2 cycle, the soil transportation, and the sequences of condensation and sublimation of the polar frosts, primarily made of both CO_2 and H_2O. At geological time scales, the surface diversity originates from a coupled evolution of the magmatic, volcanic, and alteration history of Mars, with no major role played

Fig. 9. La bocca della verita: Vassili I. Moroz (*33*).

by CO_2 within the volatile inventory: OMEGA found no definite evidence that CO_2 sustained a long-term greenhouse effect enabling liquid H_2O to remain stable at the surface of Mars in the post-Noachian terrains. The OMEGA observations to date are consistent with an early escape of most of the Mars atmosphere, leaving the atmosphere as the major present CO_2 reservoir; water has been found trapped in two sinks: as ice mixed with dust within the two large perennial polar caps and as surface hydrated minerals, which seem to have been synthesized during the early evolution of Mars.

If indeed the water remained essentially stable in solid and gaseous states during the past 3 billion years, with only transient episodes of liquid water brought to the surface, then the episodic evolution of the obliquity (*29*) might have played a major role in the observed surface composition: in condensing water ice in a variety of areas determined by the atmospheric circulation in different insolation conditions, resulting in morphological and compositional (alteration) effects, that can be observed in the present surface properties of a number of areas.

References and Notes

1. J.-P. Bibring *et al.*, *Eur. Space Agency Spec. Pub.* **1240**, 37 (2004).
2. A. Chicarro, P. Martin, R. Trautner, *Eur. Space Agency Spec. Pub.* **1240**, 3 (2004).
3. J. F. Mustard *et al.*, *Science* **307**, 1594 (2005); published online 17 February 2005 (10.1126/science.1109098).
4. The Martian chronology is described in three epochs on the basis of the density of impact craters identified on optical images. The first epoch, named Noachian, corresponds to the early heavy bombardment, which ended ~3.8 billion years ago (Noachis is a region located within the highly cratered terrains, which still cover ~40% of the Mars surface, primarily in the south). The second period, or Hesperian (named after Hesperia Planitia as a typical area of this age), lasted until ~2 billion years. It is followed by the most recent Amazonian period (Amazonia Planitia belongs to the least cratered volcanic plain).
5. J. L. Bandfield, V. E. Hamilton, P. R. Christensen, *Science* **287**, 1626 (2000).
6. J. L. Bandfield, *J. Geophys. Res.* **107**, 5092 (2002).
7. V. E. Hamilton *et al.*, *Meteorit. Planet. Sci.* **38**, 871 (2003).
8. J. F. Mustard *et al.*, *J. Geophys. Res.* **102**, 25605 (1997).
9. M. B. Wyatt, H. Y. McSween, *Nature* **417**, 263 (2002).
10. M. E. Minitti *et al.*, *J. Geophys. Res.* **107**, E5 10.1029 (2002).
11. P. H. Schultz, J. F. Mustard, *J. Geophys. Res.* **109**, E01001 (2004).
12. T. M. Hoefen *et al.*, *Science* **302**, 627 (2003).
13. F. Poulet, S. Erard, *J. Geophys. Res.* **109**, E2009 (2004).
14. J.-P. Bibring *et al.*, *Nature* **428**, 627 (2004).
15. S. Byrne, A. P. Ingersoll, *Geophys. Res. Lett.* **30**, 1696 (2003).
16. Y. Langevin *et al.*, *Science* **307**, 1581 (2005); published online 17 February 2005 (10.1126/science.1109438).
17. H. H. Kieffer, T. N. Titus, *Icarus* **154**, 162 (2001).
18. A. S. Yen, S. S. Kim, M. H. Hecht, M. S. Frant, B. Murray, *Science* **289**, 1909 (1999).
19. M. S. Milliken, J. F. Mustard, *Lunar Planet. Sci. Conf.* **XXXIV**, 1345 (2003).
20. The band depth *D* is determined by $D(\lambda) = 1 - R(\lambda)/[R(3.70) + R(2.35)]$ where *R* is the observed reflectance value at a given wavelength λ (µm). The integrated band depth is then calculated by integrating the band depth for each λ between 2.9 and 3.7 µm.
21. J.-P. Bibring, S. Erard, *Space Sci. Rev.* **96**, 293 (2001).
22. A. Gendrin *et al.*, *Science* **307**, 1587 (2005); published online 17 February 2005 (10.1126/science.1109087).
23. R. E. Arvidson *et al.*, *Science* **307**, 1591 (2005); published online 17 February 2005 (10.1126/science.1109509).
24. Y. Langevin, F. Poulet, J.-P. Bibring, B. Gondel, *Science* **307**, 1584 (2005); published online 17 February 2005 (10.1126/science.1109091).
25. S. W. Ruff *et al.*, *J. Geophys. Res.* **106**, 23921 (2001).
26. R. E. Arvidson *et al.*, *J. Geophys. Res.* **108** (E12), 8073 (2003).
27. D. T. Vaniman *et al.*, *Nature* **431**, 663 (2004).
28. Olivine weathering to sulfates has been discussed in the framework of acid fog (*30*–*32*).
29. J. Laskar *et al.*, *Icarus* **170**, 343 (2004).
30. M. Settle, *J. Geophys. Res.* **84**, 8343 (1979).
31. A. Banin *et al.*, *J. Geophys. Res.* **102**, 13341 (1997).
32. N. J. Tosca *et al.*, *J. Geophys. Res.* **109**, 10.1029/2003JE002218 (2004).
33. The OMEGA instrument has been developed with the support of Centre National d'Etudes Spatiales (CNES), Agonzia Spatiale Italiana (ASI), and Russian Space Agency. The scientific activity is funded by national space and research agencies and universities in France, Italy, Russia, Germany, and the United States of America. We are very grateful to the entire ESA teams who, together with industry, enabled this mission. We thank G. Neukum and the High Resolution Stereo Camera (HRSC) team who provided us with images of areas mapped by OMEGA. Vassili I. Moroz (Fig. 9) was a member of the OMEGA team until he passed away 23 June 2004. He was a pioneering planetologist who provided an outstanding contribution to OMEGA. This paper is dedicated to him.

REPORT

Summer Evolution of the North Polar Cap of Mars as Observed by OMEGA/Mars Express

Y. Langevin,[1]* F. Poulet,[1] J.-P. Bibring,[1] B. Schmitt,[2] S. Douté,[2] B. Gondet[1]

The Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité (OMEGA) visible-infrared imaging spectrometer extensively observed regions of Mars with latitudes above 70°N in late 2004 (heliocentric longitude from L_s 93° to L_s 127°). The extent of water ice at the surface and the size of ice grains were monitored as a function of time. Bright, small-grained frost, which initially covered a large fraction of the polar cap, waned in favor of large-grained ice. In outlying regions, dominated by large-grained ice, the albedo increased over the period. Evaluating the dust content was model dependent. However, contamination of ice by dust was low.

The permanent caps of Mars were first comprehensively observed by Viking in 1976. Thermal infrared (IR) observations by Viking and then Mars Global Surveyor (MGS) Thermal Emission Spectrometer (TES) demonstrated that the surface of the permanent north polar cap is dominated by water ice (*1*, *2*). Analyses of imaging data from Mariner 9, Viking (*3*), and MGS (*4*–*6*) have demonstrated that the albedo of the north circumpolar regions varies within a single summer season and from summer to summer. Water ice is very bright in the visible spectrum when clean, but even a small amount of dust contamination can reduce the albedo to values close to that of the dust itself if the dust grains are embedded in ice grains (*7*). Determining

[1]Institut d'Astrophysique Spatiale, CNRS–Université Paris Sud, 91405 Orsay, France. [2]Laboratoire de Planétologie de Grenoble, CNRS–Université Joseph Fourier, 38400 Grenoble, France.

*To whom correspondence should be addressed. E-mail: yves.langevin@ias.u-psud.fr

the extent of surface water ice from visible data alone is therefore not straightforward.

Northern regions of Mars at latitudes higher than 70°N could be extensively observed by the Mars Express orbiter in October to December 2004, after the summer solstice (heliocentric longitude from L_s 93° to L_s 127°; Sun elevations ranging from 35° to 22°). During this period, the OMEGA visible–near-infrared imaging spectrometer (1.2-mrad instantaneous field of view, 8.8° swath width) mapped these regions with a resolution of 2 to 5 km/pixel, a scale at which local slopes are only relevant for a few locations. The spectral range of OMEGA is 0.4 to 5.1 µm. The spectral sampling is 14 nm, and the

signal-to-noise ratio exceeds 300 from 1.1 to 2.6 µm (8, 9), a spectral range that covers the strong ice bands at 1.5 and 2 µm (H_2O) and 1.43, 2, and 2.6 µm (CO_2). The calibration uncertainty on the photometric function is ~10% and that on broad-band strengths is ~3%. Reflectance and band-strength variations of <0.5% can be monitored. Therefore, OMEGA can identify surface ices without relying on albedo. Only water ice was observed at the surface after L_s 77° (8), in agreement with (2). The strength of water ice absorption features in the 1- to 2.6-µm range strongly depends on grain size (Fig. 1). The absorption at 1.25 µm increases faster with grain size than that at 1.5 or 2 µm. For large

grains (~1 mm), the 2-µm feature is saturated, with a plateau from 1.9 to 2.15 µm. This makes it possible to monitor the extent and grain size evolution of water ice exposed on the surface both qualitatively with global maps and quantitatively by modeling the evolution of spectra for regions of interest.

Global maps were obtained by using a false-color scheme at three wavelengths (1.085, 1.242, and 1.51 µm). The red level (1.085 µm) is representative of the continuum, the blue level (1.51 µm) corresponds to a deep water ice band, and the green level (1.242 µm) corresponds to an ice band that is only strong for large grains (Fig. 1). Ice-poor regions have a similar albedo at all three wavelengths; hence, they show up as shades of gray. Small-grained surface ice (10 to 100 µm) shows up as yellow with this color scheme, because it has high albedos at 1.085 µm (red) and 1.25 µm (green), and a low albedo at 1.51 µm (blue). Large-grained surface ice (~1 mm) shows up as red in this color scheme, because it has a high albedo at 1.085 µm (red) and low albedos at 1.242 µm (green) and 1.51 µm (blue).

The first comprehensive coverage of circumpolar regions (Fig. 2) was obtained between 27 September and 8 October 2004 (L_s 93° to 98°). The distribution of surface water ice (red to yellow) is similar to that of the bright albedo features observed by Viking (3), except for a dent on the Olympia Planitia lobe. The smaller albedo notch at 205°E on this lobe observed by Mariner 9 (3) is similar to that in the OMEGA surface water ice map.

Fig. 1. Model spectra of pure water ice from 1 to 2.6 µm as a function of grain size calculated with a radiative transfer model (10). Blue: 10 µm; green: 100 µm; red: 1 mm. The dashed lines at 1.085, 1.242, and 1.51 µm correspond to the three OMEGA wavelengths that were used to monitor the evolution of grain size in the maps of Figs. 2 and 3.
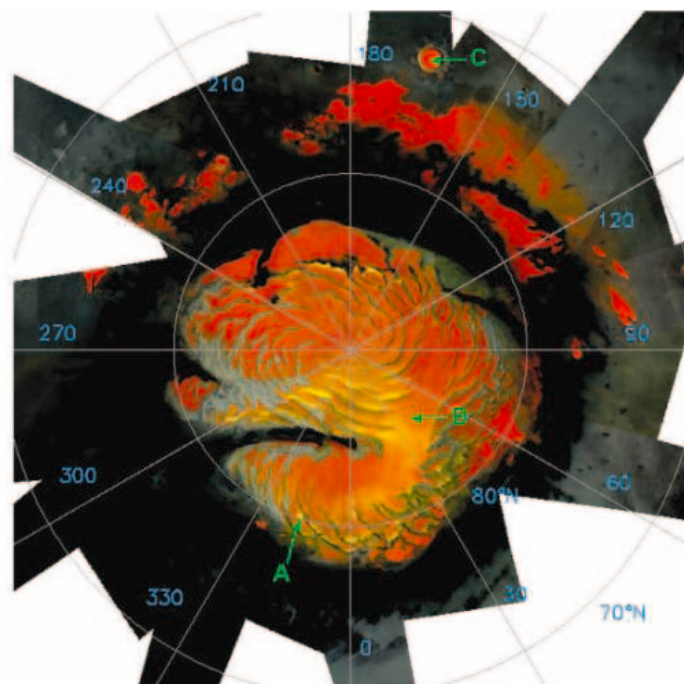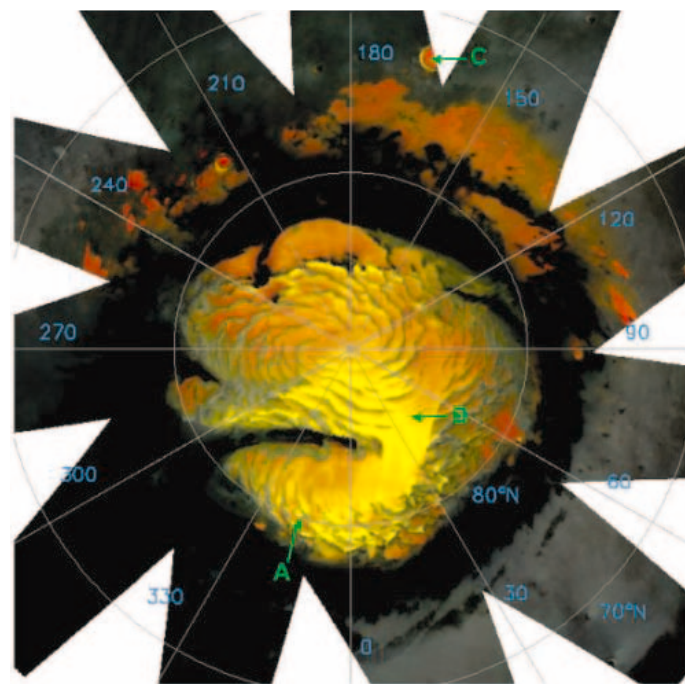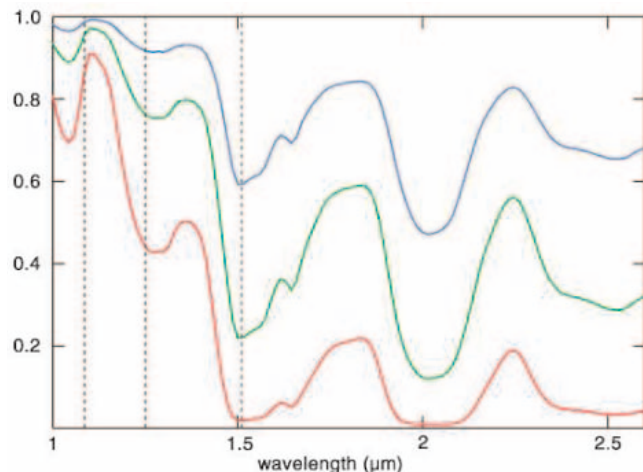
Fig. 2. (Left). RGB composite of 11 OMEGA tracks acquired between 27 September 2004 (L_s 93.3°) and 8 October 2004 (L_s 97.9°). Red color plane: Lambertian albedo at 1.085 µm; green color plane: Lambertian albedo at 1.242 µm; blue color plane: Lambertian albedo at 1.51 µm. For the each of the three color planes, the full scale (0 to 255) corresponds to an albedo range of 0.25 to 0.60. **Fig. 3 (Right).** RGB composite of nine OMEGA tracks acquired between 29 October 2004 (L_s 107.4°) and 4 November 2004 (L_s 110.3°). The RGB color scheme is the same as in Fig. 2.

Nearly half of the ice-rich regions at latitudes higher than 80°N was then covered with bright fine-grained water ice (bright yellow). Absorption strengths at 1.24 and 1.51 μm in outlying regions correspond to those of large-grained ice with a dust contribution, because the albedo contrast is weak.

A second map (Fig. 3) was obtained between 29 October and 4 November 2004 (L_s 107° to 110°). The extent of regions with exposed water ice remained very stable over the 1-month interval since the map of Fig. 2 was obtained. However, only a few small areas at the edges of the central ice-rich region are still characterized by a small grain size (bright yellow). Most regions with exposed water ice are now dominated by large grains (red). This major evolution of the grain size of water ice exposed at the surface is associated with a decrease in albedo in the continuum for regions with latitudes higher than 80°N. The albedo contrast in the continuum between ice-rich areas at latitudes lower than 80°N and the surrounding dust-covered regions is much sharper in early November than in early October, with the same spatial distribution of surface ice. Therefore, the variations in ice coverage between different martian years that have been discussed on the basis of albedo contrasts (3–6) must be considered with caution.

Indeed, as previously reported (7, 10), major albedo variations of surface ice can result from changes in grain size or from a dust contribution, with a complex interaction between these parameters. This contribution results from a combination of the following processes, which lower the apparent albedo of exposed surface ice and reduce the contrast of the water ice absorption bands.

(i) Aerosols lower the flux received by the surface (atmospheric extinction, which lowers the apparent albedo), while contributing part of the received signal by backscattering. (ii) Subpixel dust patches on the surface (areal mixing) contribute to the reflectance in proportion to their fractional coverage. (iii) An intimate mixture of dust and ice grains modifies the scattering properties of the surface layers (11, 12). This can result from cosedimentation of dust and ice grains on preexisting ice. (iv) Very small quantities of dust grains (≪1%) embedded within a water ice matrix (intra-mixture) strongly decrease the albedo of large-grained ice because they can absorb photons before their backscattering on grain boundaries (7). Such a situation can be expected for ice grains nucleated on dust seeds.

The contribution of aerosols could be assessed in early November, because it became possible with nadir pointing to observe the same area with two very different solar incidences. In particular, a region covered by bright dust at 159.7°E, 73°N was observed with solar incidences of 53.9° and 73.2°, corresponding to a factor of 2 in terms of the aerosol column density on the path of incoming solar photons. The resulting decrease in the apparent albedo at 1.085 μm was only ~3%. Therefore, atmospheric extinction by aerosols had only a minor impact on our observations at incidences lower than 60° in early November. The apparent albedo of dark areas at 1.085 μm was larger in early October than in early November (13% instead of 11.5%). This can be attributed to a 1.5% change of the aerosol backscattering contribution over 1 month, which has only a minor impact on the albedo of ice-rich regions (~50% in the continuum). Water ice absorptions lower than 10% were observed in dusty circumpolar deposits very close to surface ice; hence, ice in aerosols played a minor role in early summer, in agreement with TES observations (2).

The evolution of three regions of interest (A, B, and C in Figs. 2 and 3) was investigated in terms of grain size and dust content using a model (12) based on a radiative transfer approach (13). Regions A and B could be observed repeatedly until mid-December (L_s 127°). Region C, at a latitude of 73°, could only be observed until late October (L_s 107°). The size of water ice grains is well constrained, because absorption features are very strong (Fig. 1). However, previous modeling work (12) shows that the type of dust-ice mixing (areal mixture, intimate mixture, or intramixture) has a major influence on quantitative evaluations of the contribution of different components. We obtain an upper limit in terms of the volume fraction of dust in surface ice by first considering an intimate mixture of dust and ice grains, no surface dust patches, and no aerosol contribution.

Region A (343°E, 80.2°N) remained very bright from L_s 93° to L_s 127° (up to 70% albedo at 1.085 μm). The evolution of the observed spectra from 1 to 2.6 μm is shown on Fig. 4A. During this period of more than 2 months, the albedo at 1.085 μm slightly increased (from 0.65 to 0.7). The upper limit on the dust contamination lowered from ~3% to ~2% in volume with dark dust 5 μm in size in intimate mixture with ice grains. The best model fit is obtained with ice grain sizes lower than 200 μm, ice grains with sizes <50 μm constituting 80% of the volume. The slight reduction in contrast of the 1.65-μm feature, which is temperature dependent (14), is consistent with a temperature increasing from ~160 to ~210 K. Such a temperature evolution is within the range observed by TES at similar latitudes and values of L_s (2). The persistence of such bright patches at the edge of ice-rich regions over several months suggests that they may consist of thick accumulations of small water ice grains, equivalent to snow drifts.

Region B (42.5°E, 85.1°N) is typical of the evolution of the spectrum of permanent exposed ice on the northern ice cap (Fig. 4B). The initial spectrum (L_s 93.3°) is similar to

Fig. 4. Spectral evolution of regions of interest on the permanent cap at six values of L_s for each region. Each spectrum corresponds to the closest OMEGA pixel (2 to 5 km in size, depending on altitude), with a signal-to-noise ratio exceeding 300 from 1 to 2.6 μm. (**A**) 343°E, 80.2°N (region A in Fig. 3). (**B**) 42.5°E, 85.2°N (region B in Fig. 3). (**C**) 164.2°E, 72.4°N (Region C in Fig. 3); the spectrum at L_s 101.8° is mostly overlain by that at L_s 103.2°. The two dashed lines correspond to best-fit intimate mixture models for the initial spectrum (blue) and final spectrum (red). For region A, only the model spectrum corresponding to the final spectrum is drawn because there is no major evolution with time.
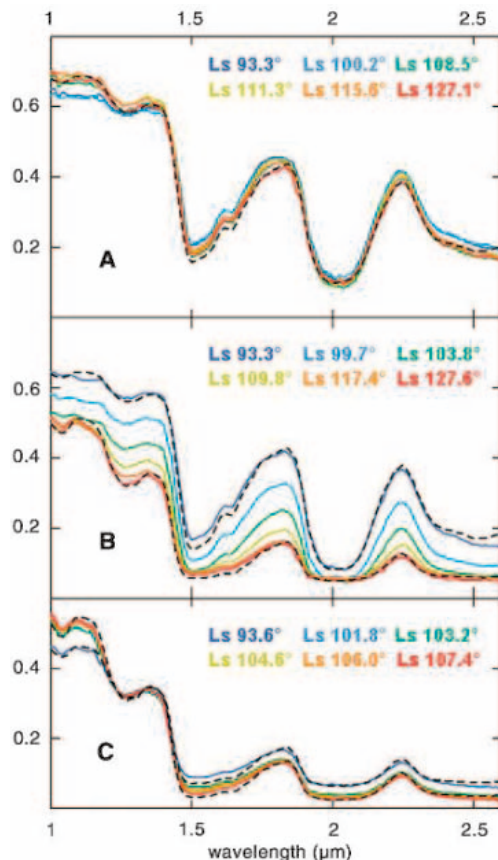
the initial spectrum of region A, dominated by ice grains in the sub–100 μm size range. The final spectrum (L$_s$ 127.6°) is characterized by a lower albedo and a saturated 2-μm absorption feature, which corresponds to a much larger ice grain size. The best model fit is obtained when ice grains with a size of 700 to 800 μm constitute 70% of the volume. Ice metamorphism cannot explain such a size change in only 1 month (7). The most likely interpretation of this spectral evolution is therefore that fine-grained frost constituting the last remnants of the seasonal cap sublimates away, in agreement with models (15, 16). This process exposes permanent ice, which is dominated by large grains (700 to 800 μm). This is similar to what is observed on terrestrial ice caps, such as Greenland, with a correlation between a decrease in albedo and the disappearance of seasonal frost (17). Dust contamination at the end of the period is 6% in terms of volume fraction, which is an upper limit because intimate mixture is assumed.

Region C lies within Korolev, one of the southernmost ice-filled craters, at a latitude of 73°N. From L$_s$ 93.6° to L$_s$ 107.4° (Fig. 4C), the spectra are characterized by a flat-bottomed absorption feature at 2 μm; hence, most of the volume (60 to 70%) is occupied by ice grains with sizes in the 700-μm to 1-mm range, similar to the final spectrum of region B. At L$_s$ 93.6°, the observed spectrum has an albedo at 1.085 μm of 46% and an albedo at 2 μm of 6%; hence, the best model fit indicates an initial dust contamination (9% assuming intimate mixture) that is slightly larger than that in region B at the end of its evolution (6%). At L$_s$ 107.4°, the albedo at 1.085 μm has increased to 52%, after which spectral evolution slows

down. The albedo at 2 μm of the final spectrum is only 2.5%. Therefore, the upper limit of intimate dust contamination is low (<5%).

If part of the dust is embedded in the ice matrix (intramixture), spectral modeling of the observations leads to a negligible volume fraction of dust (≪1% in all cases), in agreement with (9), and the estimated grain size remains the same as with the intimate mixture. If one also takes into account a possible contribution of aerosols, the conclusion is that the residual large-grained ice that is exposed since early summer in outlying regions and 1 month later on most of the ice cap itself is likely to be very clean. Either there is very little dust deposition during the global dust storm season (contrary to what is observed at lower latitudes) or there is an effective clean-up process during sublimation of the surface layers. The decrease in dust contribution that is observed between L$_s$ 93° and L$_s$ 107° in outlying regions (Fig. 4C) supports a surface clean-up process, even if a decrease in the optical thickness of aerosols cannot be totally excluded.

We have shown that in the central part of the north permanent cap of Mars, albedo variations in early summer are linked to a major increase in ice grain size, as seasonal frost with grain sizes <100 μm sublimates, so that the larger grains of the permanent ice (~1 mm) dominate the reflectance properties. In a few areas close to the edges of permanent surface ice regions, bright accumulations of small grains survive until late in the summer. In outlying ice-rich regions, the albedo markedly increases whereas the grain size (also ~1 mm) does not change. One month after the summer solstice, old ice is

exposed at the surface over most ice-rich areas. This should lead to a net sublimation of ice with the present inclination. The low level of dust contamination of the ice remains to be explained, although there are indications that a clean-up process may be associated with sublimation.

References and Notes
1. H. H. Kieffer, S. C. Chase, T. Z. Martin, R. E. Miner, F. D. Palluconi, *Science* **194**, 1341 (1976).
2. H. H. Kieffer, T. N. Titus, *Icarus* **154**, 162 (2001).
3. D. S. Bass, K. E. Herkenhoff, D. E. Paige, *Icarus* **144**, 382 (2000).
4. P. B. James, B. A. Cantor, *Icarus* **154**, 131 (2001).
5. A. S. Hale, D. S. Bass, L. K. Tamppari, *Lunar Planet. Sci.* **34**, 1422 (2003).
6. M. C. Malin, K. S. Edgett, *J. Geophys. Res.* **106**, 23429 (2001).
7. H. H. Kieffer, *J. Geophys. Res.* **95**, 1481 (1990).
8. J.-P. Bibring *et al.*, *Science* **307**, 1576 (2005); published online 17 February 2005 (10.1126/science.1108806).
9. The visible part of the range would be useful for constraining surface dust content and composition. However, the visible channel (0.4 to 1 μm) still has some calibration problems, and atmospheric aerosols have a major influence at short wavelengths.
10. W. M. Calvin, T. N. Titus, *Lunar. Planet. Sci.* **35**, 1455 (2004).
11. S. Douté, B. Schmitt, *J. Geophys. Res.* **103**, 31367 (1998).
12. F. Poulet, J. N. Cuzzi, D. P. Cruikshank, T. Roush, C. M. Dalle Ore, *Icarus* **160**, 313 (2002).
13. Y. Shkuratov, L. Starukhina, H. Hoffmann, G. Arnold, *Icarus* **137**, 235 (1999).
14. W. M. Grundy, B. Schmitt, *J. Geophys. Res.* **103**, 25809 (1998).
15. M. A. Mischna, M. I. Richardson, R. J. Wilson, D. J. McCLeese, *J. Geophys. Res.* **108**, 5062 (2003).
16. B. Levrard, F. Forget, F. Montmessin, J. Laskar, *Nature* **431**, 1072 (2004).
17. A. W. Nolin, *J. Geophys. Res.* **103**, 25851 (1998).
18. This work was supported by CNES, CNRS, and Université Paris XI.

REPORT

Sulfates in the North Polar Region of Mars Detected by OMEGA/Mars Express

Yves Langevin,* François Poulet, Jean-Pierre Bibring, Brigitte Gondet

The Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité (OMEGA) imaging spectrometer observed the northern circumpolar regions of Mars at a resolution of a few kilometers. An extended region at 240°E, 85°N, with an area of 60 kilometers by 200 kilometers, exhibits absorptions at wavelengths of 1.45, 1.75, 1.94, 2.22, 2.26, and 2.48 micrometers. These signatures can be unambiguously attributed to calcium-rich sulfates, most likely gypsum. This region corresponds to the dark longitudinal dunes of Olympia Planitia. These observations reveal that water alteration played a major role in the formation of the constituting minerals of northern circumpolar terrains.

The northern circumpolar regions of Mars constitute a complex geological region. Permanent ice deposits partially cover a bulge attributed to a water ice cap, as well as some

low-altitude outlying regions down to latitudes of ~70°N. Intermixed with these bright areas, layered terrains and dark circumpolar sand deposits are observed. This entire region is a

young complex system of layered deposits of ice and dust, plains, and dunes (1). The role of liquid water in the origin of these complex structures has been debated, either as outflows (2) or as large standing bodies of water (3, 4). Mineralogical information is needed to constrain the geological history of these regions. Few unambiguous signatures have been observed in the northern circumpolar regions

Institut d'Astrophysique Spatiale (IAS), Bâtiment 121, 91405 Orsay Campus, France.

*To whom correspondence should be addressed. E-mail: yves.langevin@ias.u-psud.fr

by previous experiments dedicated to mineralogical studies (5, 6).

The OMEGA visible and near-infrared imaging spectrometer has been operating around Mars since January 2004. This instrument maps 352 wavelengths from 0.35 to 5.1 μm (7, 8). A series of 11 observations obtained between 13 and 24 October 2004, corresponding to heliocentric longitude L_s 100° to 105° (northern summer), provides comprehensive coverage of the regions above 73°N. A composite of the 11 observations (Fig. 1) combines information on the reflectance in the continuum at 1.085 μm (gray scale) with the strength of an absorption feature at 1.927 μm. An irregular crown of low-albedo regions corresponding to the dark circumpolar dunes is observed around the regions dominated by water ice and layered deposits. The absorption at 1.927 μm is defined relative to a continuum at 1.857 μm and 2.136 μm. These three wavelengths have been selected so as to lie out of the main atmospheric CO_2 absorption bands. The regions with exposed ice (9) are left in white. A region where the absorption at 1.927 μm is larger than 20% (orange-red area in Fig. 1) is identified close to 240°E, 80°N. The extent of this region, which has a mean reflectance of 16% at 1.08 μm, is ~60 km by 140 km. This spectral signature is observed on only a small fraction of the dark circumpolar dunes (Fig. 1). The corresponding spectral unit lies close to the outer boundary of the permanent surface ice.

A representative reflectance spectrum from a 10 km by 10 km spectrally homogeneous region centered at 244.5°E, 80.2°N is shown in Fig. 2. This region lies 4.9 km below the Mars Orbiter Laser Altimeter (MOLA) reference surface; hence, strong atmospheric absorption features are observed, in particular at 2 μm. As a first approximation, atmospheric absorption features can be removed by dividing the reflectance spectrum by a ratio of spectra of homogeneous terrains obtained by OMEGA on Olympus Mons, scaled to the same column density of CO_2. The reflectance spectrum corrected from atmospheric absorption (red spectrum in Fig. 2) exhibits a strong absorption feature at ~1.94 μm and weaker signatures at other wavelengths. The 1.94-μm absorption feature is far stronger (>25%) than the closest atmospheric H_2O feature at 1.875 μm (<2%). Spectra from most other dark circumpolar areas are characterized by a much weaker and broader absorption feature also centered close to 1.94 μm. When mapping the strength of the 1.94-μm absorption feature in Fig. 1, we did not select the OMEGA spectral channel closest to maximum absorption (1.941 μm) but instead chose that at the next lower wavelength (1.927 μm), because atmospheric CO_2 absorption is much smaller at 1.927 μm (~2%) than at 1.941 μm (~10%). The spatial distribution of weaker features (e.g., at 2.48 μm and

1.75 μm) is highly consistent with that of the 1.94-μm absorption feature, which demonstrates that a single constituent is responsible for these spectral signatures over the whole region.

We applied a spectral ratio method so as to consolidate our mineral identification. We divided the blue spectrum from Fig. 2 (before atmospheric correction) by the reflectance spectrum of a region with a similar albedo and altitude that did not exhibit a narrow signature. The spectral ratio (red curve in Fig. 3) does not depend on knowledge of the OMEGA photometric function. The main atmospheric bands have a small impact on this spectral ratio, because the altitudes of the two regions are similar (4.9 km and 4.4 km, respectively, below the MOLA reference surface), except for possible spatial variations in water vapor content (at most a few percent in band strength). The spectral features that are observed in the normalized ratio are similar to

that of the reflectance spectrum of Fig. 2, which shows that the observed strong signature at 1.94 μm is not an artifact resulting from the OMEGA photometric function or the atmospheric correction procedure.

In Fig. 3, we also show the ratio of two spectra obtained by OMEGA during ground calibration: that of a gypsum powder with grain sizes < 40 μm and that of an aluminum oxide powder, which is spectrally featureless in this wavelength range. The match in terms of position and relative strength of the minor and major features between the Mars spectral ratio and the calibration spectral ratio constitutes an unambiguous detection of gypsum ($CaSO_4 \cdot 2H_2O$) as a major constituent of the hydrated region observed close to 245°E, 80°N. Less hydrated calcium sulfates, such as bassanite ($2CaSO_4 \cdot H_2O$), would also provide a good match with the observed spectral signature, but the fit is definitely less satisfactory, in particular for the small spectral
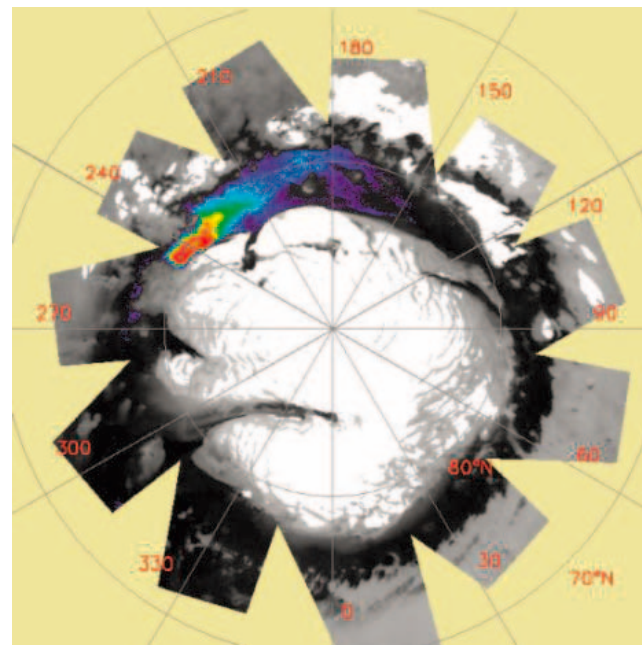
Fig. 1. Composite map of 11 OMEGA observations obtained in October 2004. The gray scale corresponds to the albedo at 1.085 μm (continuum), with a range from black (10%) to light gray (35%). False-color information is superimposed for areas exhibiting an absorption feature at 1.927 μm, with a rainbow color scale from 6% (purple) to >25% (red) in terms of band strength. White areas correspond to the regions where OMEGA observes water ice at the surface, which are defined by a water ice band strength at 1.5 μm exceeding 20% (9).
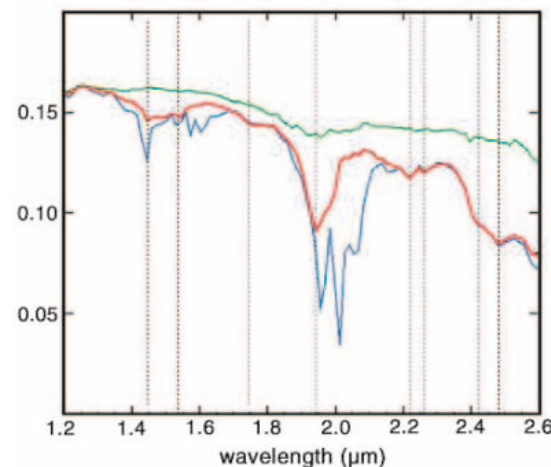
Fig. 2. Reflectance spectra. Blue curve, reflectance spectrum of the region exhibiting the strongest absorption at 1.94 μm (244.5°E, 80.2°N); red curve, blue spectrum corrected for atmospheric absorption; green curve, reference spectrum of a dark region of similar elevation and albedo (54.9°E, 78.2°N) that does not exhibit a strong and narrow signature at 1.94 μm, corrected for atmospheric absorption. Absorption features are observed at 1.445, 1.535, 1.745, 1.94, 2.22, 2.26, 2.42, and 2.48 μm (dotted lines) in the red spectrum.

features observed at 1.445, 1.535, 2.22, 2.26, and 2.42 μm. Fine-grained gypsum is bright (reflectance >40% in the continuum). Therefore, the observed reflectance of 16% at 1.1 μm requires an admixture of opaque and spectrally neutral material such as magnetite.

The gypsum signature is spatially correlated to dark longitudinal dunes (*1*) that belong to Olympia Planitia, the largest sand sea on Mars. Both physical and morphological properties led to the idea that the Olympia dunes were formed at their present location: MOLA topography reveals that most of Olympia Planitia has a convex topography and appears to underlie the present polar cap and associated deposits (*10*). Furthermore, the Olympia dune field has a much lower thermal inertia than other Martian dunes, which supports the idea that it consists of aggregates of μm-sized particles (*11, 12*). Such small particles could be easily transported via atmospheric suspension, and other sources for dune material have been proposed, such as pyroclastic ash and oceanic or outflow sediment (*10*). However, transportation of dust and sand from distant sources in the present low-pressure atmosphere is unlikely (*12*). Byrne and Murray (*13*) have identified a platy unit distinct from the layered deposits at the base of the present layered deposits. The regions where this unit is exposed at the surface are correlated with the occurrences of dune bed forms, leading these authors to the conclusion that the dune material is derived from in situ erosion of the platy unit.

The calcium sulfate component observed by OMEGA supports the hypothesis that the constitutive material of the dunes was formed in a different environment than the other polar terrains, most likely before the emplacement of the present-day polar cap (*13, 14*). The formation of calcium sulfates requires an interaction between water and pyroxenes or feldspars in a sulfur-rich environment. Volcanic processes provide the most likely supply of sulfur (H_2S, SO_2, or sulfur-rich pyroclastic ashes). Several processes can be considered

for the formation of sulfates, involving surface weathering, groundwater circulation in buried sediments, acid snow or rain falling on basalt, or evaporation of standing bodies of water.

An increase in planetary obliquity may have increased the quantity of water vapor in the atmosphere and the abundance of water ice clouds in the polar region (*15*), so that circumpolar terrains may have undergone strong atmospheric weathering. However, the alteration signature observed in bright terrains is quite different from that of the sulfate-rich region, with no absorption at either 1.75 μm or 1.94 μm. Furthermore, the supply of sulfur is not straightforward in this hypothesis. The formation of sulfate minerals on Mars could also occur through oxidative weathering of iron sulfides (*16, 17*). The prerequisite for such a weathering scenario is a wide distribution of sulfides. If this weathering process has occurred on Mars in the past, large quantities of sulfates should have been formed. This would require the weathered material to be exposed on only a small fraction of the planet, where OMEGA can detect it (*8*).

Groundwater related to hydrothermal sources or local volcanism is not a likely formation process for the observed sulfate-rich unit, as no clear geological evidence has been found for volcanic activity in circumpolar regions. Apart from local geothermal hot spots, other scenarios such as basal melting at high pressure under thick deposits of dust and ice have been proposed as possible explanations for groundwater-related processes [e.g., (*18*)].

The interaction of basalts with acidic snow has been proposed as a candidate for the formation of sulfates in Noachian terrains (*8, 19*). Such a process requires extended episodes of volcanic activity. The small extent of the sulfate-rich unit is difficult to reconcile with a volcanic episode during which the whole region would have been exposed to acidic snows, unless this unit represents a small fraction of a much larger area that was recently uncovered by erosion.

The observed sulfate-rich unit could also result from outflows from an ice cap during a warm climatic excursion. The sulfates would then result from the evaporation of water saturated with salts after prolonged interactions with sulfur-rich material and crustal rocks. Such hypersaline conditions would extend the stability zone of the liquid phase near the surface to colder temperatures. If this interpretation is correct, our results provide strong evidence for climatic episodes compatible with the presence of surface water, as advocated by Pathare and Paige (*20*) during periods of high obliquity (>45°).

The mineralogical signature of gypsum observed by OMEGA in a specific region close to the permanent north polar cap provides strong evidence for alteration processes involving water on the surface of Mars. Possible hypotheses for the formation of this unit involve the interaction of acidic snow with calcium-rich minerals during an extensive episode of volcanic activity, major outflows followed by evaporation of salt-rich water (possibly predating the formation of the present-day ice cap), or a combination of these two processes.

References and Notes
1. K. Tanaka, D. Scott, *USGS Misc. Invest. Ser. Map I-1802-C* (1987).
2. V. R. Baker, M. Carr, V. Gulick, C. Williams, M. Marley, in *Mars*, H. Kieffer et al., Eds. (Univ. of Arizona Press, Tucson, AZ, 1992), pp. 493–522.
3. T. Parker, R. Saunders, D. Schneeberger, *Icarus* **82**, 111 (1989).
4. S. Clifford, T. Parker, *Lunar Planet. Sci. Conf.* **30**, abstract 1619 (1999).
5. M. B. Wyatt, K. L. Tanaka, *3rd International Conference on Mars Polar Science and Exploration* (Lake Louise, Alberta, Canada, 13 to 17 October 2003), abstract 8118.
6. S. Ivanov et al., *6th International Conference on Mars* (Pasadena, CA, 20 to 25 July 2003), abstract 3182.
7. J.-P. Bibring et al., *ESA Spec. Pub. 1240*, 37 (2004).
8. J.-P. Bibring et al., *Science* **307**, 1576 (2005).
9. Y. Langevin et al., *Science* **307**, 1581 (2005); published online 17 February 2005 (10.1126/science.1109438).
10. K. E. Fishbaugh, J. W. Head, *J. Geophys. Res.* **105**, 22455 (2000).
11. D. A. Paige, J. E. Bachman, K. D. Keegan, *J. Geophys. Res.* **99**, 25959 (1994).
12. K. E. Herkenhoff, A. R. Vasavada, *J. Geophys. Res.* **104**, 16487 (1999).
13. S. Byrne, B. C. Murray, *J. Geophys. Res.* **107**, 10.1029/2001JE001615 (2002).
14. K. E. Fishbaugh, J. W. Head III, *Lunar Planet. Sci. Conf.* **25**, abstract 1156 (2004).
15. M. A. Mischna, M. I. Richardson, R. J. Wilson, D. J. McGleese, *J. Geophys. Res.* **108**, 10.1029/2003JE002051 (2003).
16. R. G. Burns, D. S. Fisher, *J. Geophys. Res.* **95**, 14415 (1990).
17. J. L. Bishop, P. Schiffman, M. D. Lane, M. D. Dyar, *Int. J. Astrobiol.*, in press.
18. S. Clifford et al., *Icarus* **144**, 210 (2000).
19. A. Gendrin et al., *Science* **307**, 1587 (2005).
20. A. V. Pathare, D. A. Paige, *1st International Conference on Mars Polar Science and Exploration* (Camp Allen, TX, 18 to 22 October 1998), abstract 953.
21. Supported by Centre National d'Etudes Spatiales, CNRS, and Université Paris Sud.

Fig. 3. Normalized spectral ratios. Red curve, ratio of the blue spectrum in Fig. 2 divided by the raw reflectance spectrum of the reference region (corresponding to the green spectrum in Fig. 2 without the atmospheric correction), normalized to 1 at 1.25 μm; blue curve, normalized ratio of the spectra of gypsum and aluminum oxide obtained by OMEGA during ground calibration. Features are observed in both spectral ratios at 1.445, 1.535, 1.745, 1.94, 2.22, 2.26, 2.42, and 2.48 μm (dotted lines).
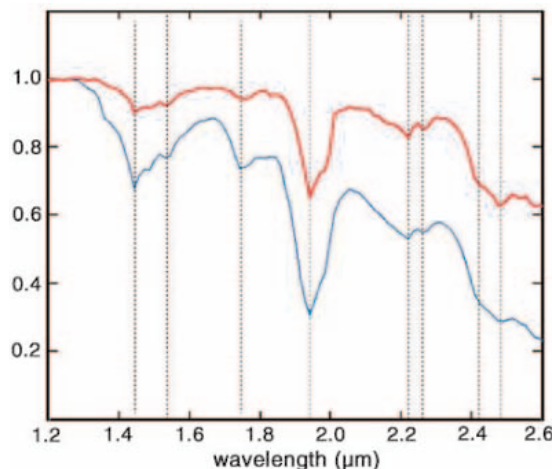
Sulfates in Martian Layered Terrains: The OMEGA/Mars Express View

Aline Gendrin,[1]* Nicolas Mangold,[2] Jean-Pierre Bibring,[1] Yves Langevin,[1] Brigitte Gondet,[1] François Poulet,[1] Guillaume Bonello,[1] Cathy Quantin,[3] John Mustard,[4] Ray Arvidson,[5] Stéphane LeMouélic[6]

The OMEGA/Mars Express hyperspectral imager identified hydrated sulfates on light-toned layered terrains on Mars. Outcrops in Valles Marineris, Margaritifer Sinus, and Terra Meridiani show evidence for kieserite, gypsum, and polyhydrated sulfates. This identification has its basis in vibrational absorptions between 1.3 and 2.5 micrometers. These minerals constitute direct records of the past aqueous activity on Mars.

OMEGA, a visible–near infrared hyperspectral imager (0.35 to 5.1 μm) on board the European Space Agency/Mars Express mission (1), has been performing the surface mapping at a km to sub-km resolution since January 2004 (2). Here, we present the identification and mapping of hydrated sulfates found in geological units exhibiting light-toned layered terrains.

We used wavelengths of 1.3 to 2.5 μm to identify sulfate minerals. In this wavelength range, solar reflected light dominates the spectrum, and the planetary thermal emission is negligible. Consequently, the retrieval of surface and atmospheric contributions is straightforward (3). Although most of the fundamentals of the relevant vibrational absorption signatures lie outside this domain, their overtones and combinations are diagnostic and lead to unambiguous mineralogical identifications through the accurate and combined determination of band position, shape, and depth.

In this wavelength range, the OH and H_2O vibration features dominate. Most hydrated minerals have absorption bands at ~1.4 and ~1.9 μm (Fig. 1). These bands are caused by overtones and combinations of the symmetric H–O–H stretch (v_1), the asymmetric H–O–H stretch (v_3), and the H–O–H bend (v_2) (4). From 2.2 to 2.4 μm, narrow absorptions due to metal–O–H vibrations (4) are present and are diagnostic of mineralogy. S–O bonds also create absorption features in this range when associated with H_2O (Table 1). Spe-

cifically, the $(SO_4)^{2-}$ stretch, with fundamentals near 10 μm, produces overtones from 2.2 to 2.5 μm. Their precise position and shape vary with composition and result in features at 2.2 and 2.4 μm for hydrated sulfates.

The identification in the OMEGA data set of specific combinations of the central position, shape, and depth of absorption bands, with reference to calibration spectra performed on terrestrial analogs, enables us to characterize three spectral types attributed to sulfates as follows.

The first spectral type is characterized by absorptions at 1.6, 2.1, and 2.4 μm (5). The 1.6- and 2.1-μm absorptions correspond to the 1.4-μm and 1.9-μm features shifted to longer wavelengths. Such a shift is observed only in monohydrated sulfates (Fig. 1). It is
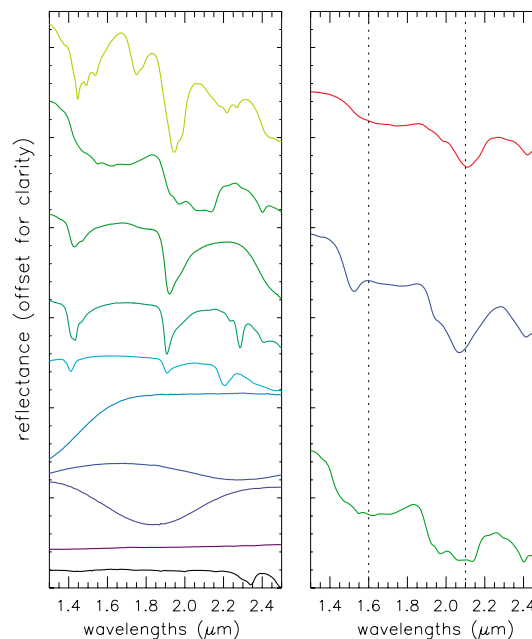
Fig. 1. Laboratory spectra of terrestrial analogs (31). (**Left**) From bottom to top: calcite (carbonate), hematite (oxide), enstatite (orthopyroxene), diopside (clinopyroxene), olivine, montmorillonite (phyllosilicates), nontronite (phyllosilicate), chabazite (zeolite), kieserite (sulfate), and gypsum (sulfate). (**Right**) Monohydrated sulfates, from bottom to top: $MgSO_4 \cdot H_2O$ (kieserite), $MnSO_4 \cdot H_2O$, and $ZnSO_4 \cdot H_2O$.

[1]Institut d'Astrophysique Spatiale (IAS), Bâtiment 121, 91405 Orsay Campus, France. [2]Interactions et Dynamique des Environnements de Surface (IDES), Bâtiment 509, 91405 Orsay Campus, France. [3]Laboratoire de Sciences de la Terre, Bâtiment géode, 69100 Villeurbanne, France. [4]Geological Sciences, Brown University, Providence, RI 02912, USA. [5]Earth and Planetary Sciences, Washington University, St. Louis, MO 63130, USA. [6]Planétologie, Université de Nantes, 44322 Nantes, France.

*To whom correspondence should be addressed. E-mail: Aline_Gendrin@brown.edu

Table 1. Spectral types identified in the OMEGA data set, corresponding best candidate mineral, and geographic location. Band assignments have their basis in previous studies (29, 30).

Detected absorption bands, position (μm)	Corresponding best candidate mineral	Geographic location
~1.6,* ~2.1,* 2.4†	Kieserite	Ius, Hebes, Capri, Candor, Melas, Juventae Chasma, Aram Chaos, Terra Meridiani
~1.4,* ~1.9,* 2.4† drop	Polyhydrated sulfate	Ius, Hebes, Capri, Candor, Melas, Ophir Chasma, Terra Meridiani
~1.4,* ~1.75, ~1.9,* ~2.2 (2.21‡ and 2.27‡), 2.4† drop	Gypsum	Juventae Chasma, Iani Chaos

*Combinations and overtones of OH⁻ or H_2O bending and stretching fundamentals.　　†$3v_3(SO_4)^{2-}$, H_2O overtone and combinations.　　‡$v + \delta$ (H_2O)

caused by the strong coupling between the single water molecule and the sulfate ion (*6*). In hydrated environments, the distinct 2.4-μm band is caused by the $(SO_4)^{2-}$ stretch. These features are diagnostic of monohydrated sulfates. Mg monohydrated sulfate, kieserite ($MgSO_4 \cdot H_2O$), matches the OMEGA spectrum over the 1.4- to 2.5-μm wavelength range (Fig. 2).

The second spectral type is characterized by five absorption bands at 1.4, 1.75, 1.9, 2.2, and 2.4 μm (Fig. 2). The 1.4- and 1.9-μm absorption bands indicate a hydrated mineral. Two humps are observed inside the 1.4-μm band (*7*). The 2.2-μm absorption band also has a double shape, with two small absorptions at 2.21 and 2.27 μm. The 2.2- and 2.4-μm features are caused by the $(SO_4)^{2-}$ stretch in hydrated environments. This set of absorptions matches the absorptions observed in calcium sulfates, and gypsum ($CaSO_4 \cdot 2H_2O$) or bassanite ($2CaSO_4 \cdot H_2O$) are the best candidates.

The third spectral type has 1.4- and 1.9-μm features characteristic of hydration and a 2.4-μm feature due to the $(SO_4)^{2-}$ stretch in hydrated environment. The features have asymmetrical shapes exhibiting plateau-like right wings. These are diagnostic of polyhydrated sulfates, and sulfates with different cations provide good matches to the OMEGA spectrum. Epsomite ($MgSO_4 \cdot 7H_2O$) fits well the OMEGA spectrum, as does copiapite [$Fe^{2+}Fe_4^{3+}(SO_4)_6 (OH)_2 \cdot 20H_2O$] or halotrichite [$Fe^{2+}Al_2(SO_4)_4 \cdot 22H_2O$] (*6, 8*).

These sulfate minerals (*9*) are associated with light-toned layered deposits (*10–14*)
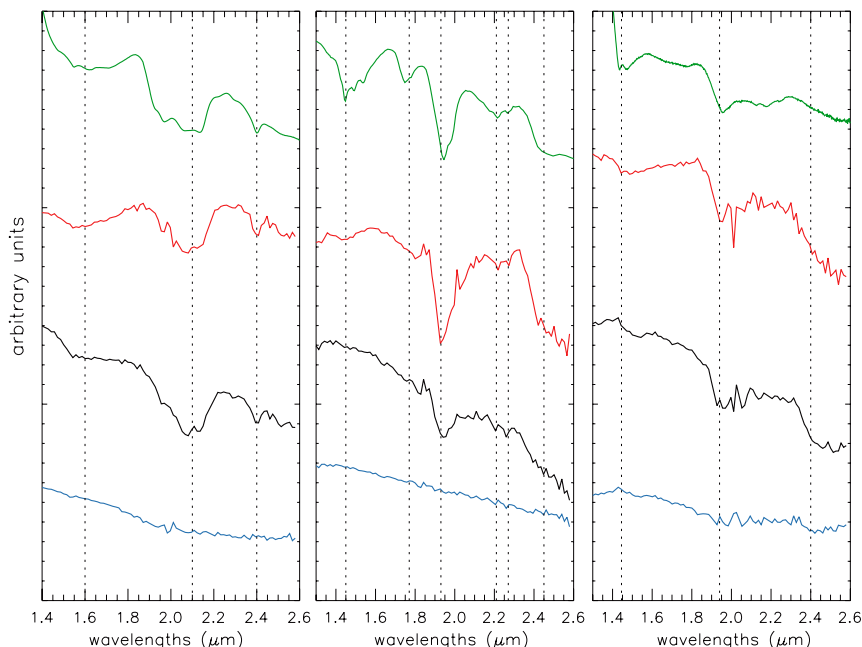
Fig. 2. Sulfate spectra identified in the OMEGA observations. Green, library spectrum; black, OMEGA spectrum; blue, reference spectrum; red, spectral ratio. (**Left**) Kieserite $MgSO_4 \cdot H_2O$(28). (**Center**) Gypsum. (**Right**) Polyhydrated sulfate. The library spectrum corresponds to epsomite ($MgSO_4 \cdot 7H_2O$), but other polyhydrated sulfates are good analogs in this spectral domain.

Fig. 3. Location of the deposits identified in Valles Marineris and surroundings plotted over the MOLA altimetry. (**A**) Red, kieserite type; green, polyhydrated sulfate; blue, spectral class most likely corresponding to gypsum; pink, other hydrated minerals, identified by their absorptions at 1.4 and 1.9 μm. Sulfates are identified over light-toned deposits in (**B**) Ius Chasma floor (orbit 71), (**C**) East Hebes Chasma interior (orbit 360), (**D**) Juventae Chasma interior (orbit 460), Capri Chasma (**E**) crater flank (orbit 61) and (**F**) interior buttes (orbit 363), (**G**) Iani Chaos (orbit 353), and (**H**) Aram Chaos interiors (orbit 401).
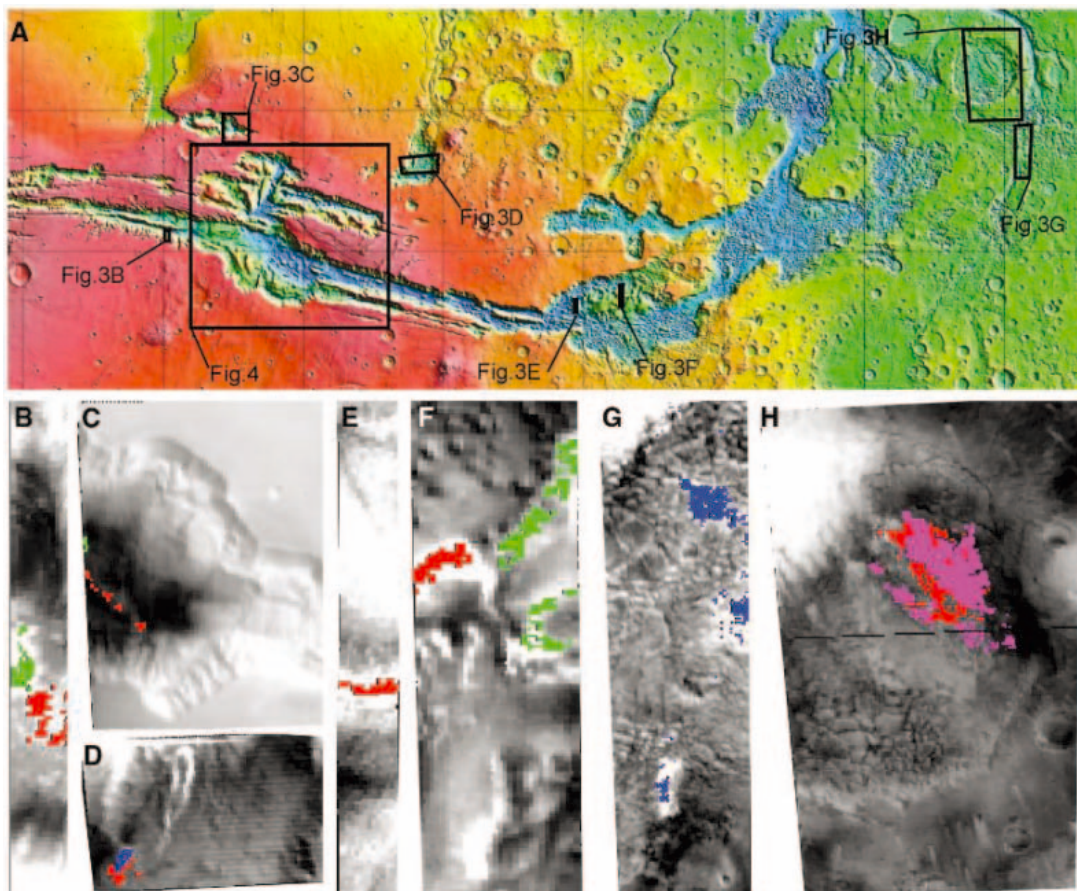
in Valles Marineris, Margaritifer Terra, and Terra Meridiani (Figs. 3 to 6), described as follows.

Fig. 4. Same color code as in Fig. 3 for Melas, Candor, and Ophir Chasmata. We combined orbits numbered 308, 334, 360, 515, 548, and 581 to obtain the mosaic.
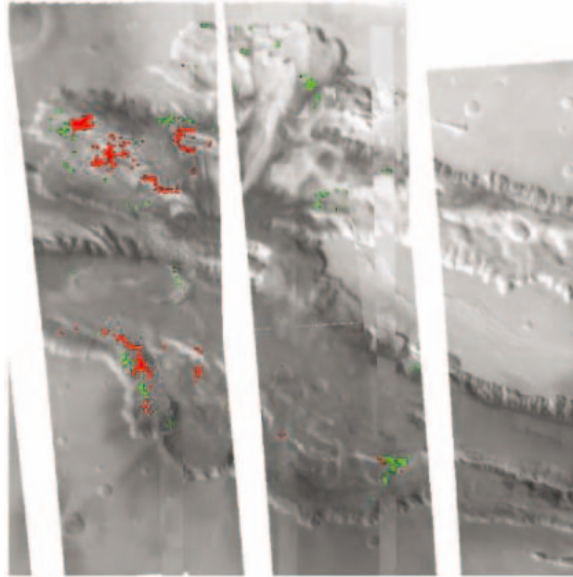
In the main canyons of Valles Marineris, kieserite and polyhydrated sulfates are associated with light-toned deposits (Figs. 3

and 4). Mars Orbiter Camera onboard Mars Global Surveyor images of Candor and Melas indicate that these are layered terrains. In Candor Chasma, kieserite is observed on three sides of a 4-km elevated hill (Fig. 5A); the darker parts of the mesa, corresponding to eolian mantling, have no sulfate signatures. In Melas Chasma, the small buttes at the foot of Valles Marineris flanks display layering; they fit the OMEGA pixels in which polyhydrated sulfates are detected (Fig. 5B). The deposits of Juventae Chasma (Fig. 5C) are peculiar, because gypsum is identified within the horizontal layers some hundred meters in thickness each. Kieserite surrounds gypsum and is associated with a rougher material at lower elevation. The kieserite-rich area slightly extends by a few pixels over the sand dunes surrounding the layered deposits.

In Margaritifer Terra, remnants of layered deposits are scattered in different chaos locations (*14*, *15*). Several deposits inside Iani Chaos (Fig. 3G) appear to be gypsum-rich, with a good correlation with layers observed on high-resolution images (Fig. 5D). Aram Chaos

Fig. 5. Examples of correlation between sulfate deposits identified by OMEGA and layered deposits from Mars Orbiter Camera (MOC) images (Malin Space Science Systems/Jet Propulsion Laboratory/NASA). (**A1** to **D1**) OMEGA classification, see Fig. 3 for color code. (**A2** to **D2**) MOC context images, except (**B2**), THEMIS daytime mosaic. (**A3**) to (**D3**) MOC high-resolution images: (A) Candor Chasma, MOC number E03-01293 (74.5°W, 5.2°S). (B) Melas Chasma, MOC number R09-02972 (74.5°W, 11.8°W). (C) Juventae Chasma, MOC number E23-01035 (62.6°W, 4.5°W). (D) Iani Chaos, MOC number M14-01230, 18.8°W, 4.3°S.
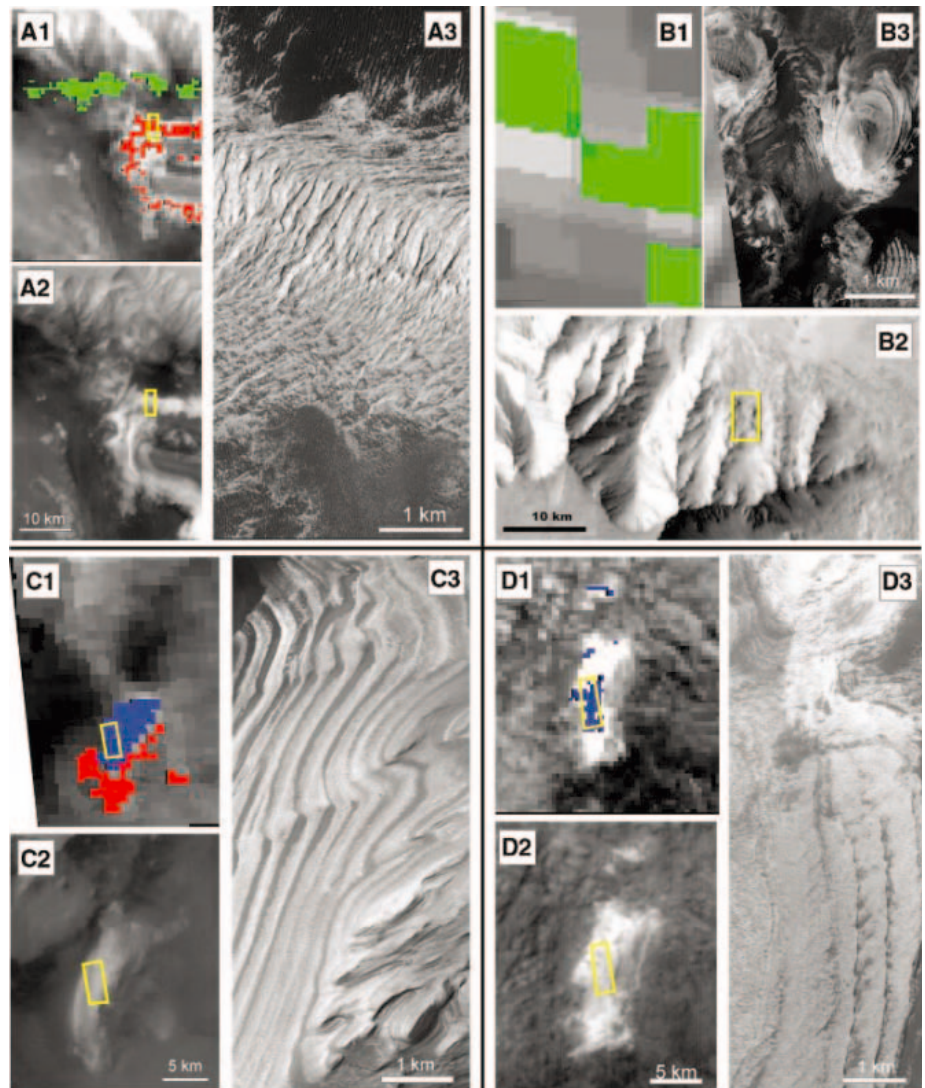
shows a higher spectral diversity (Fig. 3H) where kieserite surrounds other hydrated minerals.

In Terra Meridiani (Fig. 6), sulfates are found associated to a variety of hydrated minerals in different locations over large areas, constituting the so-called etched units, which are composed of light-toned deposits (16). In one specific area, sulfates are abundant (Fig. 6), with kieserite mapped in red and hydrated minerals probably corresponding to polyhydrated sulfates in green. The sulfate-rich unit underlies the crystalline hematite one explored by the Mars Exploration Rover Opportunity: Some of the layers discovered by Opportunity would correspond to the uppermost unit of the etched terrains (17) in which OMEGA locally detects sulfates, indicating that these sulfates are widespread and not restricted to the landing site area.

Not all Martian layered deposits observed by OMEGA exhibit absorption bands typical of hydrated minerals or sulfate minerals. Minerals without spectral signatures in this range, such as halite, might be present in the observed areas.

The stability of kieserite requires specific thermodynamical conditions. Experiments (18) show that kieserite is easily converted to hexahydrite ($MgSO_4\cdot6H_2O$) and epsomite ($MgSO_4\cdot7H_2O$) when exposed to water. However, upon further desiccation, these minerals are not transformed back to kieserite. Therefore, kieserite is not likely to be formed and preserved in deposits that have experienced surface cycles of hydration and desiccation.

The sulfates are a bulk component of the deposits, rather than a soil surface coating.

Magnesium sulfates are known to constitute a substantial component (about 5%) of Martian soils (19). OMEGA observed sulfates in the equatorial regions exclusively in close relation with light-toned layered deposits and not scattered in the surrounding soil. Moreover, most sulfate-rich deposits correspond to fresh exhumed surfaces, as seen by flutes and yardangs typical of eolian erosion, by the lack of small (<100 m) impact craters, and by the thermal signature from THEMIS/Odyssey (19). The formation of these layered deposits is dated to the Noachian and/or Hesperian epochs (12, 13).

By analogy with terrestrial geology, it has been proposed that sulfates would originate from evaporation of standing bodies of water such as shallow lakes (playas) or lagunas connected to the seawater (sabkhas) (20). Large accumulations of gypsum in the terrestrial rock record are common, whereas other sulfates such as kieserite generally exist in much lower amounts (21). However, it is likely that martian brines are more enriched in Fe and Mg relative to terrestrial standards, if they result from the alteration of the mafic crust, leading to a higher content of Mg sulfates and acidic conditions (22, 23) favoring the formation of sulfates relative to carbonates and other salts (24). The formation of evaporites on Mars would require stable liquid water in lakes or seas, which severely constrains the atmospheric properties, even considering that brines can be liquid at temperatures well below 273 K. If deposits in Valles Marineris are interpreted as lacustrine, one must account for sulfates observed over a wide range of elevations (from –4 km

in Ius Chasma to +3 km in Candor Chasma) and not merely as playa deposits in the current bottoms of canyons. It could imply that sediments were formed during the opening of the Valles Marineris canyon and accumulated in shallow water bodies like in terrestrial rift zones (10–13). In Terra Meridiani, the sulfate-rich unit is observed close to the Opportunity landing site, where parts of the layers are interpreted as subaqueous sediments (25).

Sulfates may also form at depths in the presence of sulfur-rich fluids in hydrothermal contexts (20, 26). Layered deposits could have originated as volcanic ash or eolian deposits and could have been altered later to sulfates by acidic groundwater circulation. Possibly hydrothermal conditions would also have favored sulfate deposition by subsurface brine seepage. Such a scenario could also have occurred during the rifting of Valles Marineris (10), favoring sulfur-rich fluid circulation and the alteration of volcanic ashes into sulfates, as well as in chaotic terrains (15) and Terra Meridiani (27). Lastly, sulfates could result from alteration of mafic minerals by rain and/or frost made acidic through volcanic outgassing (2).

References and Notes

1. J.-P. Bibring et al., European Space Agency Spec. Publ. **1240**, 37 (2004).
2. J.-P. Bibring et al., Science **307**, 1576 (2005); published online 17 February 2005 (10.1126/science.1108806).
3. J. F. Mustard et al., Science **307**, 1594 (2005); published online 17 February 2005 (10.1126/science.1109098).
4. R. N. Clark et al., J. Geophys. Res. **95**, 12653 (1990).
5. Band positions are given with an accuracy of 0.1 µm when the absorption bands are broad or their center is in a wavelength range with possible atmospheric residuals, or in a wavelength range with possible remaining instrumental effects. Otherwise, positions are given with an accuracy of 0.01 µm.
6. J. K. Crowley et al., Geochem. Explor. Environ. Anal. **3**, 219 (2003).
7. Y. Langevin et al., Science **307**, 1584 (2005); published online 17 February 2005 (10.1126/science.1109091).
8. M. Lane et al., Geophys. Res. Lett. **31**, L19702 (2004).
9. The sulfates are identified with the use of band ratios. To map the sulfate areas, we used a methodology based on (28). We then checked that each spectrum of the resulting classification presents the absorption bands of interest.
10. J. F. McCauley et al., Icarus **17**, 289 (1972).
11. B. K. Lucchitta et al., NASA Tech. Memo. **85127**, 233 (1982).
12. B. K. Lucchitta et al., J. Geophys. Res. **99**, 3783 (1994).
13. S. S. Nedell et al., Icarus **70**, 409 (1987).
14. M. C. Malin, K. S. Edgett, Science **290**, 1927 (2000).
15. M. G. Chapman, K. L. Tanaka, Icarus **155**, 324 (2002).
16. R. E. Arvidson et al., J. Geophys. Res. **108**, ROV 14-1 (2003).
17. B. M. Hynek, Nature **431**, 156 (2004).
18. D. T. Vaniman et al., Nature **431**, 663 (2004).
19. A. Banin et al., J. Geophys. Res. **102**, 13341 (1997).
20. J. Warren, Evaporites: Their Evolution and Economics (Blackwell Science, Oxford, UK, 1999).
21. Kieserite forms at temperatures typically higher than 40°C, whereas more hydrated phases such as epsomite form at lower temperatures. The association of these minerals of different hydration levels is common on Earth because of diurnal or seasonal temperature variations. Ions and cations present in the body of water control the sequence of formation of evaporitic minerals. The classic terrestrial sequence is dominated by large proportions of halite
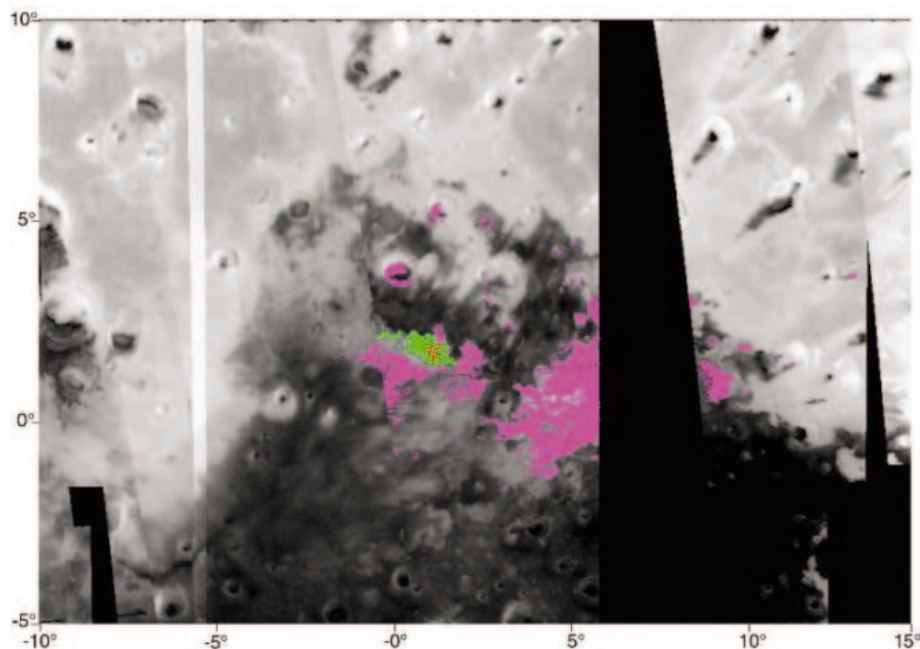
Fig. 6. Same color code as in Fig. 3 for Terra Meridiani. The mosaic is obtained combining orbits 171, 278, 291, 314, 397, 430, 485, 518, 529, and 551.

with underlying gypsum and minor deposits of carbonates at the base. Kieserite and other K or Mg salts are minor components appearing at the top of the sequence. We can outline the large accumulation of kieserite of the Zechstein basin of the Permian epoch and the deposition of magnesium sulfates in the currently drying Aral Sea.

22. M. A. Bullock et al., Icarus 170, 404 (2004).
23. D. C. Catling, J. Geophys. Res. 104, 16453 (1999).
24. A. G. Fairen et al., Nature 431, 423 (2004).
25. S. W. Squyres et al., Science 306, 1698 (2004).
26. D. L. Huston et al., Earth Planet. Sci. Lett. 220, 41 (2004).
27. R. E. Arvidson et al., Science 307, 1591 (2005); published online 17 February 2005 (10.1126/science.119509).
28. Green et al., IEEE Trans. Geosc. Rem. Sens. 26, 65 (1993).
29. G. R. Hunt et al., Mod. Geol. 3, 1 (1971).
30. J. Bishop, E. Murad, Am. Miner., in press.
31. Laboratory reflectance data provided through a cooperation with NASA and the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) team. Reflectance measurements were acquired at the Reflectance Experiment Laboratory (RELAB) at Brown University. Efforts to acquire these data by J. F. Mustard are gratefully acknowledged.

REPORT

Spectral Reflectance and Morphologic Correlations in Eastern Terra Meridiani, Mars

R. E. Arvidson,[1]* F. Poulet,[2] J.-P. Bibring,[2] M. Wolff,[3] A. Gendrin,[2] R. V. Morris,[4] J. J. Freeman,[1] Y. Langevin,[2] N. Mangold,[2] G. Bellucci[5]

The Mars Express Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité (OMEGA) hyperspectral image data covering eastern Terra Meridiani indicate the ubiquitous presence of molecular water in etched terrain materials that disconformally overlie heavily cratered terrains and underlie the hematite-bearing plains explored by the Opportunity rover. Identification of crystalline water in kieserite ($MgSO_4 \cdot H_2O$) is linked to materials exposed in a valley and plateau to the north of hematite-bearing plains. The mineralogical similarities between the etched terrain deposits examined with OMEGA data and the layered rocks examined by Opportunity imply that the ancient aqueous environments inferred from analyses of the rover data extend over regional scales.

The Mars Express instrument, OMEGA, has acquired hyperspectral imaging data of the martian atmosphere and surface since January 2004 (1). In this Report, we summarize analyses of OMEGA data from Orbit 485 over Terra Meridiani, acquired on 7 June 2004 at approximately 09:00 Mars local time and a solar incidence angle of ~46°. This Report complements the work on identification and mapping of high-latitude hydrated sulfate minerals (2) and in equatorial to mid-latitude layered deposits (3) from OMEGA data. Our focus is the identification of hydrated phases associated with etched terrain materials exposed to the east and north of the hematite-bearing plains that define Meridiani Planum (4–6).

The geologic setting of Terra Meridiani is among the most complex on Mars and includes exposures of Noachian-aged cratered terrains that have been dissected by channel systems that were carved by water flowing from the southeast to the northwest (7) (Fig. 1). After

channeling ceased, several hundred meters of layered materials were disconformally deposited onto the dissected cratered terrains. To the north of Terra Meridiani, the deposits were then buried by the southern edge of a dust mantle that peaks in thickness over Terra Arabia (5). Subsequent wind erosion then exposed the layered deposits in Terra Meridiani and differentially shaped these materials into a set of domes, ridges, plateaus, and other landforms collectively termed etched terrains (5). The hematite-bearing Meridiani Planum surfaces and materials mapped from Mars Global Surveyor (MGS) Thermal Emission Spectra (TES) (8) and explored by the Mars Exploration Rover, Opportunity (9), are at the top of the stratigraphic section of preserved layered materials. Analyses of rover-based data show that the hematite spherules and associated fragments have been concentrated on the surface as a lag deposit as wind erosion has eroded the weak spherule-bearing, sulfate-rich rocks that underlie the plains (10).

OMEGA Orbit 485 data cover portions of the eastern hematite-bearing plains, sections of etched terrains exposed to the east and north of the plains, and extensive areas of dissected and mantled cratered terrains (Fig. 1). Furthermore, the northern exposures of etched terrains include deposits in a valley and plateau oriented to the northwest-southeast and located just to the north of the transition from

the hematite-bearing plains to the etched terrains (Figs. 1 and 2). After retrieval of surface Lambert normal albedos (11), standard reduction analyses using principal components techniques (12) were used to extract spectral endmembers. Extractions focused on the wavelength interval from 1.0 to 2.6 μm because interactive inspection of spectra suggested that this region showed the greatest variety of spectral shapes. Five spectral endmembers (Fig. 3) explain 86% of the variance for plains, etched terrains, and surrounding dark cratered terrains: (i) a "bright etched terrain" spectrum typical of the signatures for etched terrain, (ii) a "dark etched plateau" spectrum representative of darker signatures associated with the northwest–southeast–trending valley and plateau (Figs. 1 and 2), (iii) a "bright plains" spectrum, (iv) a "dark plains" spectrum, and (v) a "dark crater floor" spectrum (located in the cratered terrain).

Between 0.4 and 1.0 μm, each of the five endmembers is characterized by an absorption edge between 0.4 and ~0.6 μm, a slope change near 0.6 μm, a relative reflectance maximum at ~0.75 μm, and a band minimum at longer wavelengths. The position of the minimum ranges from ~0.82 μm (feature A, Fig. 3) for the bright etched terrain spectrum to ~1.0 μm (feature B in Fig. 3) for the dark crater floor endmember. The spectral slope between ~0.6 and ~1.5 μm is steep and positive for the bright etched terrain spectrum and slightly negative for the dark crater floor spectrum. All of these spectral features have been detected in previous observations of Mars made at lower spatial resolution as compared to OMEGA data (~2 km for Orbit 485 data versus ~20 km) (13, 14). The spectral features for the bright etched terrain spectrum are consistent with the presence of Fe^{3+}-bearing minerals, interpreted to be a mixture of nanophase ferric iron and hematite (15),

[1]Department of Earth and Planetary Sciences, Washington University, St. Louis, MO 63130, USA. [2]Institut d'Astrophysique Spatiale, Batiment 121, 91405, Orsay Campus, France. [3]Space Science Institute, Boulder, CO 80301, USA. [4]NASA Johnson Space Center, Houston, TX 77058, USA. [5]Istituto Nazionale Di Astrofisica Istituto di Fisica dello Spazio Interplanetario, Via Fosso del Cavaliere, 00133 Rome, Italy.

*To whom correspondence should be addressed. E-mail: arvidson@wunder.wustl.edu

although the minimum for well-crystalline and chemically pure hematite is more typically 0.85 to 0.87 μm (*16*). On the other hand, the dark crater floor spectrum is interpreted to result from the presence of relatively unaltered materials with a thin Fe^{3+}-bearing mineral alteration or dust-rich surface layer. This spectrum clearly shows evidence for the Fe^{2+} electronic transition absorption bands (features B and C at ~1 and 2 μm, respectively, in Fig. 3) that are associated with pyroxene in basaltic materials (*17*).

The bright plains spectral endmember, on the basis of examination of MGS Mars Orbital Camera and Odyssey Thermal Emission Imaging System (THEMIS) image data, shows that the material represented by this endmember is located where hematite-bearing, mottled plains dominate the surface. In some locations, layering is exposed along relatively steep slopes. Additionally, the bright plains spectrum has a similar shape but lower overall reflectance as compared with the bright etched terrain endmember (Fig. 3). We interpret the spectral similarity between bright plains and etched terrain endmembers, and the appearance of the bright plains, as indicative of wind erosion and exposure of some etched terrain material in these areas, with a partial cover of hematite-bearing spherules and basaltic sands similar to what was found at the Opportunity site several hundred kilometers to the southwest (*18*). The partial cover subdues the water-related spectral features typical of well-exposed etched terrains. The dark plains endmember is spectrally flat from ~1.3 to 2.6 μm and is representative of the spectral signatures for slightly rolling, smooth plains with homogeneous brightness. Bowl-shaped craters are present and well preserved. No exposed layering is evident. These dark plains are interpreted to be covered with an areally uniform dark lag deposit of hematite spherules and basalt sands formed as the plains were slowly eroded by wind. Based on mapping the areal abundance of hematite from TES data (*5*), the bright and dark plains imaged by OMEGA during Orbit 485 have higher basalt sand–to–hematite spherule abundances as compared with the plains examined by Opportunity.

The deep band at ~3 μm in all spectra (feature D in Fig. 3) results from the well-known stretching fundamental vibrations of the H_2O molecule (v_1 and v_3) (*19*). This feature has been previously detected, including over Terra Meridiani (*20*). It is estimated that <4% water is needed to produce the magnitude of the observed absorptions for Mars (*21*), although the enhanced depths for the bright etched terrain as compared with those of other regions suggest a larger number. The feature at 1.92 μm (feature E in Fig. 3) is well expressed in the spectra for bright etched terrain and is diagnostic of the presence of the

Fig. 1. (**A**) Shaded relief map of Meridiani Planum region showing the dissected cratered terrain (DCT) disconformably overlain by etched terrain deposits (ET), which in turn are covered by hematite-bearing plains (PH). To the north, the deposits are mantled by younger deposits (MCT). The Opportunity landing site is shown, along with boxes that delineate an enlarged portion of the shaded relief map (red) shown in (B) and coverage for Orbit 485 OMEGA data (blue) shown in (C). Units are updated from (*5*). Topographic information is from MGS Mars Orbital Laser Altimeter (MOLA) data (*27*). Simulated sun is from west with an angle of 15° above horizon. (**B**) Shaded relief map shown in (A) with color-coded elevations overlain and location of THEMIS image subframe (C) denoted by black box. Elevation ranges from a high of –690 m (red) to a low of –1750 m (purple) relative to the International Astronomical Union–2000 coordinate system and MOLA-based areoid. (**C**) OMEGA false-color infrared composite (band wavelengths 1.2992, 1.7576, and 2.3947 μm as blue, green, and red, respectively) covering eastern Terra Meridiani with locations of shaded relief map and THEMIS coverage [(A) and (B)] shown as red and black boxes, respectively. Locations for which OMEGA spectra endmember spectra were extracted are also shown.
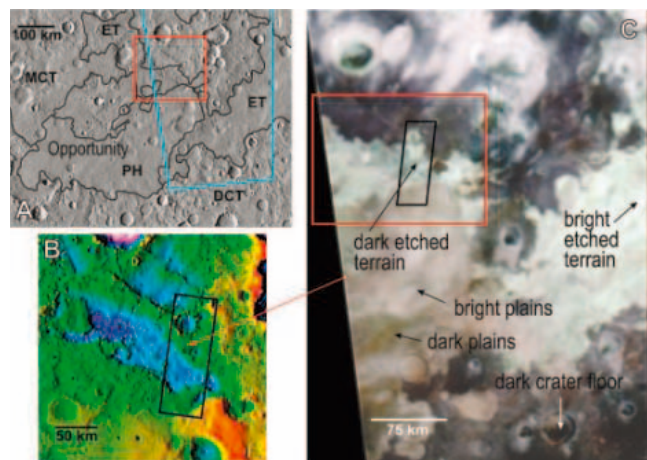
Fig. 2. (**A**) THEMIS scaled daytime infrared image subframe showing major geomorphic units including etched terrain materials exposed on a plateau and in a valley. White line shows location of vertical profile shown in (C). Image ID: I02434005BTR.img acquired on 2 July 2002 at ~16:00 Mars local time. Regions covered in the image all have elevated and similar thermal inertias (*5*). Thus, the overall reflectance in the visible and reflected infrared controls the afternoon temperature. Warm surfaces (high brightness) in the THEMIS data correspond to the dark (and thus absorbing in solar wavelengths) regions in the OMEGA data shown in Fig. 1C. (**B**) Enlarged view of portion of THEMIS subframe showing the valley and plateau. The dark etched endmember was derived from the plateau region. (**C**) Vertical topographic profile from MOLA data (128 pixels/degree bins) with 100 times vertical exaggeration.
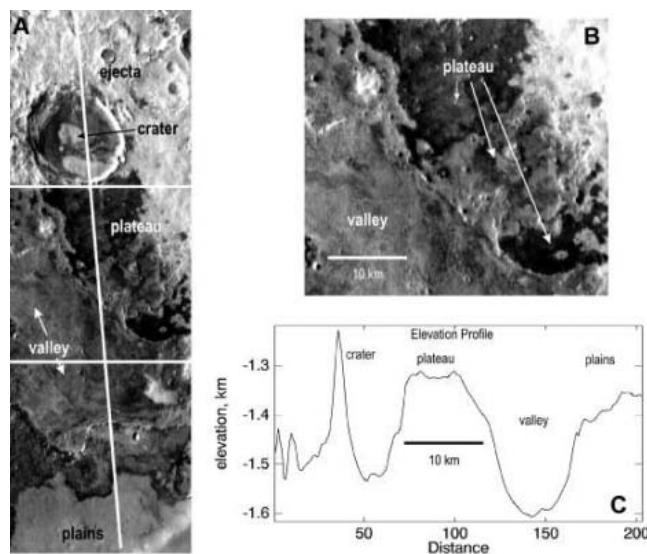
H_2O molecule and results from combination of two fundamental vibrations ($v_2 + v_3$) (*19*). In the absence of spectral evidence for hydrated and hydroxolated phases in the spectrum, we associate feature E with water adsorbed onto surface materials. The strength of the ~3- and 1.92-μm features for etched terrain spectra implies that materials exposed in these deposits are enhanced in their water contents relative to the other units for which spectral endmembers were extracted. A spectral feature near 1.4 μm should also be present for these water-bearing materials but is not

apparent (Fig. 3), presumably because it is too weak to be observed in the retrieved reflectance data.

The last three spectral features, corresponding to F (~2.5 μm) in the bright etched terrain spectrum and to G (~2.1 μm) and H (~2.4 μm) in the dark etched plateau spectrum, require additional analyses before assignments can be made. Each spectrum was normalized to the endmember spectrum for the dark plains for the 1.6- to 2.6-μm region. The dark plains spectrum is flat in this wavelength interval and thus provides a reason-

Fig. 3. OMEGA-based reflectance spectra for spectral endmember units with letters to denote spectral features. Gaps correspond to wavelengths with channels exhibiting anomalous behavior or not properly modeled with our radiative transfer techniques. The vertical offsets at ~1 μm for some spectra are associated with residual calibration errors.
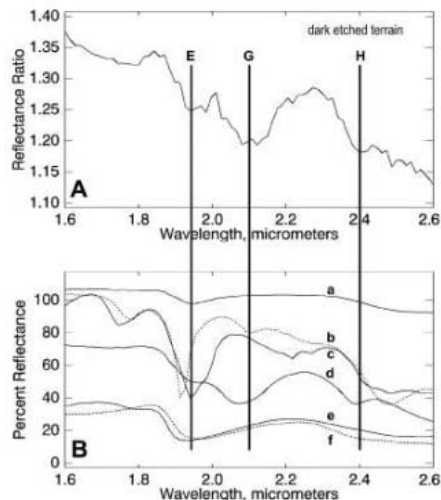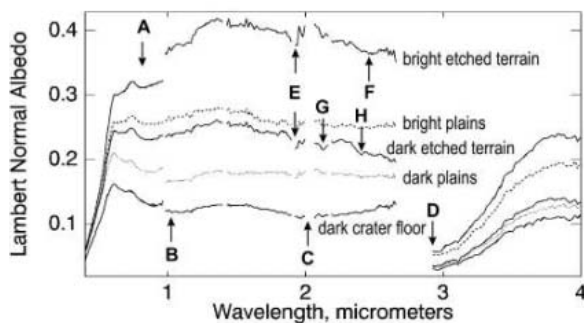
Fig. 4. (A) OMEGA-based reflectance spectrum for dark etched plateau region divided by the spectrum for dark plains and compared in (B) with laboratory spectra for various sulfate minerals. Minerals: a, barite ($BaSO_4$ with a trace amount of water); b, bassanite ($CaSO_4 \cdot 0.5H_2O$); c, gypsum ($CaSO_4 \cdot 2H_2O$); d, kieserite ($MgSO_4 \cdot H_2O$); e, calcium fluoride (CaF_2) mixed with a small amount of water to simulate water adsorbed on mineral surfaces; and f, epsomite ($MgSO_4 \cdot 7H_2O$). For each sample, sand-sized particles were run in a diffuse reflectance mode in a Nicolet 670 Fourier Transform Spectrometer system fitted with KBr optics, HgCdTe-B detector, and Harrick Cricket (TM) reflectance accessory. Reflectance is expressed relative to a diffuse gold surface.

able spectrum for normalization. The reason for the normalization is to minimize artifacts resulting from imperfect instrument calibration and atmospheric removal. The ratio spectrum enhances the depths of these features G and H and the 1.92-μm H_2O feature (Fig. 4). The ratio spectrum was compared with a number of laboratory-based reflectance spectra of a range of minerals, focusing on sulfate-bearing minerals that have water of crystallization after reconnaissance comparisons suggested that hydrated sulfates provide the best matches (Fig. 4). The bands at ~2.1 and ~2.4 μm, which are interpreted to be combination bands involving the stretching vibration of H_2O molecules within a crystal lattice, best match our laboratory spectrum of kieserite ($MgSO_4 \cdot H_2O$;

Fig. 4). The spectrum of the monohydrate of ferrous sulfate (szomolnokite) is similar in this region, but the presence of the mineral is not indicated because its ferrous band minima near 0.94 and 1.33 μm and reflectance maximum near 0.67 μm (22, 23) are not evident in OMEGA spectra. A similar ratio operation for the bright etched terrain spectrum produced a strong band (feature E) consistent with the presence of adsorbed molecular water as discussed above. The identification of H_2O-bearing minerals in the etched terrains, and the presence of kieserite for the valley and plateau, does not mean that the relevant materials are composed only of these phases. Rather, these mineral phases are present in relatively high abundances relative to other areas we examined in this paper. On the basis of our analyses, etched units are also enriched in Fe^{3+}-bearing minerals. In addition, modeling of TES spectral emissivity data suggests that pyroxene, plagioclase feldspar, and glass are also present in etched terrain deposits (5).

Comparison of OMEGA and THEMIS data shows that the dark etched plateau region with the kieserite signature is located between two deposits of the more typical bright etched terrain (Fig. 2). Furthermore, the kieserite-bearing surface exhibits a set of domes that is not typical of the rest of the valley and plateau system (Fig. 2). These domes also appear in the valley on the left-hand edge of the THEMIS frame shown in Fig. 2, exactly where OMEGA data also show a spectrum indicative of kieserite. Perhaps the domes evident in the valley and plateau and associated with a kieserite signature are preserved eruptive volcanic cones that were buried, altered by corrosive ground water, and then exhumed by wind. Alternatively, the domes could be a consequence of preferential cementation of materials by ground water preferentially flowing along fractures. Subsequent wind erosion may have left these regions as local domes because of their indurated nature. Examination of THEMIS data for all the etched terrains evident in Orbit 485 suggests that preferential cementation along fractures, followed by differential erosion to produce ridges and plateaus, was a common occurrence (24).

The most important conclusion of this paper is that the etched terrain layered materials exhibit evidence for molecular water adsorbed onto the surface materials and, for the dark etched plateau, incorporated as structural water in the hydrated sulfate mineral kieserite. These results complement the analyses conducted by the Opportunity rover on the bright layered rocks in Eagle and Endurance craters, where evidence for hydrated sulfate minerals was also found, including kieserite (25). It is our conclusion that Opportunity examined the upper section of the etched terrain deposits in these craters and that our results imply that aqueous processes were involved in forming and/or altering the etched terrain materials over distances of hundreds of kilometers and throughout the several-hundred-meter thickness of the etched terrain deposits.

References and Notes

1. J.-P. Bibring et al., Science 307, 1576 (2005); published online 17 February 2005 (10.1126/science.1108806).
2. Y. Langevin et al., Science 307, 1584 (2005); published online 17 February 2005 (10.1126/science.1109091).
3. A. Gendrin et al., Science 307, 1587 (2005); published online 17 February 2005 (10.1126/science.1109087).
4. B. M. Hynek et al., J. Geophys. Res. 107, 5088, 10.1029/2002JE001891 (2002).
5. R. E. Arvidson et al., J. Geophys. Res. 108, 8073, 10.1029/2002JE002041 (2003).
6. B. M. Hynek, Nature 431, 156 (2004).
7. B. M. Hynek, R. J. Phillips, Geology 29, 407 (2001).
8. P. R. Christensen et al., J. Geophys. Res. 105, 9623, 10.1029/1999JE001093 (2000).
9. S. W. Squyres et al., Science 306, 1698 (2004).
10. R. E. Arvidson et al., Science 306, 1730 (2004).
11. OMEGA data for Orbit 485 over Terra Meridiani were reduced to Lambert normal albedos for each of the 352 wavelengths through the use of a multiple scattering code that models atmospheric aerosol (and Rayleigh) scattering and molecular absorption (26) with atmospheric parameters derived from Opportunity and TES data acquired close in time to data for Orbit 485. For wavelengths longer than 2.5 μm, additional terms in the model included thermal emission of the surface and atmosphere. Examination of surface reflectance retrievals demonstrates that atmospheric features are properly modeled and/or the radiometric calibrations for the instrument are well understood for all but the longest wavelengths in the OMEGA data (4.00 to 5.08 μm). Consequently, OMEGA data from this wavelength interval are not discussed in this Report.
12. A. A. Green et al., IEEE Trans. Geosci. Remote Sensing 26, 65 (1988).
13. J. F. Mustard, J. F. Bell III, Geophys. Res. Lett. 21, 353 (1994).
14. S. Murchie et al., Icarus 147, 444 (2000).
15. R. V. Morris et al., J. Geophys. Res. 105, 1757 (2000).
16. R. V. Morris et al., J. Geophys. Res. 90, 3126 (1985).
17. J. F. Mustard et al., Science 307, 1594 (2005); published online 17 February 2005 (10.1126/science.1109098).
18. J. F. Bell et al., Science 306, 1703 (2004).
19. G. Herzberg, Molecular Spectra and Molecular Structure (Van Nostrand Reinhold Company, New York, 1945).
20. A. M. Baldridge, W. M. Calvin, J. Geophys. Res. 109, E04S90, 10.1029/2003JE002066 (2004).
21. A. S. Yen et al., J. Geophys. Res. 103, 11125 (1998).
22. J. K. Crowley et al., Geochem. Explor. Environ. Anal. 3, 219 (2003).
23. J. L. Bishop et al., in preparation.
24. For an example, see THEMIS image V03445003, which shows areas within the eastern etched terrain exhibit a polygonal ground pattern. In places, the polygons have been eroded to expose interconnected ridges consisting of materials emplaced within the valleys separating the polygons.
25. P. R. Christensen et al., Science 306, 1733 (2004).

26. K. Stamnes *et al.*, *Appl. Opt.* **27**, 2502 (1988).
27. D. E. Smith *et al.*, *J. Geophys. Res.* **106**, 23689, 10.1029/2000JE001364 (2001).
28. R.E.A., M.W., and R.V.M. acknowledge NASA support

REPORT

Olivine and Pyroxene Diversity in the Crust of Mars

J. F. Mustard,[1]* F. Poulet,[2] A. Gendrin,[2] J.-P. Bibring,[2] Y. Langevin,[2] B. Gondet,[2] N. Mangold,[3] G. Bellucci,[4] F. Altieri[4]

Data from the Observatoire pour la Minéralogie, l'Eau, les Glaces, et l'Activité (OMEGA) on the Mars Express spacecraft identify the distinct mafic, rock-forming minerals olivine, low-calcium pyroxene (LCP), and high-calcium pyroxene (HCP) on the surface of Mars. Olivine- and HCP-rich regions are found in deposits that span the age range of geologic units. However, LCP-rich regions are found only in the ancient Noachian-aged units, which suggests that melts for these deposits were derived from a mantle depleted in aluminum and calcium. Extended dark regions in the northern plains exhibit no evidence of strong mafic absorptions or absorptions due to hydrated materials.

The igneous composition of the martian crust has been examined through remotely sensed data, meteorites, and in situ observations by landers and rovers (*1*). Meteorites exhibit the greatest petrologic diversity but are, with the exception of one sample, <1.3 billion years in age and thus young. Remotely sensed and landed measurements imply that, where exposed, the igneous crust is dominantly basaltic, composed mostly of feldspar and pyroxene (*2*, *3*). Two major divisions in crustal composition are recognized on the basis of their thermal infrared (IR) spectral signatures (*2*). Type I material, predominantly in the equatorial highlands, is basaltic. Type II, found predominantly in the northern lowland plains, has been interpreted to be andesite or basaltic andesite (*4*) as altered basalt with a large component of hydrolytic weathering materials (*5*, *6*), oxidized basalt (*7*), or silica-coated basalt (*8*). There have been a few outcrops of ancient crust identified in thermal emission data that exhibit concentrations of olivine (*9*) and low-calcium pyroxene (LCP) above the limits of detection (*10*).

Here, we present the first results for the crustal composition of Mars derived from the OMEGA reflectance observations. Visible/near-infrared (NIR) reflectance measurements are most sensitive to the presence of iron-bearing mafic minerals. These analyses complement existing observations, help to

resolve issues, and provide insight into the crustal composition and evolution.

The OMEGA experiment and operations are described elsewhere (*11*, *12*). This analysis focuses on visible/NIR reflectance measurements of OMEGA (0.35 to 2.6 μm), where observations of surface reflectance also include atmospheric contributions from dust, water ice aerosols, CO_2, and H_2O vapor. We perform an atmospheric correction assuming that the surface and atmospheric contributions are multiplicative and that the atmospheric contribution follows a power law variation with altitude (*13*). An atmospheric spectrum is derived from a high-resolution observation crossing the summit of Olympus Mons. Assuming a constant surface contribution, the ratio of a spectrum from the base of Olympus Mons to one over the summit provides the atmospheric spectrum at a power function of their difference in altitude. The atmospheric contribution to each spectrum is then removed by dividing the observation by the derived atmospheric spectrum, scaled by the strength of the CO_2 atmospheric absorption measured in the observation.

Olivine and pyroxene are two important classes of rock-forming minerals that have absorption bands in the visible/NIR that result from electronic crystal field transitions of Fe in octahedral coordination (*14*). These absorptions are diagnostic of the minerals and their chemical composition (*15*, *16*). Olivine [$(Mg,Fe)_2SiO_4$] has a broad, complex absorption centered near 1 μm that varies in width, position, and shape with increasing Fe content (*16*). Pyroxenes [$(Ca,Fe,Mg)_2Si_2O_6$] are recognized by the presence of two distinct absorptions centered near 1 and 2 μm, where

the band centers shift toward longer wavelengths with increasing calcium content. LCPs (e.g., orthopyroxene) have short-wavelength band centers (0.9 and 1.8 μm), whereas high-calcium pyroxenes (HCPs) (e.g., clinopyroxene) typically have long-wavelength band centers (1.05 and 2.3 μm) (*15*). Laboratory measurements of these minerals and their mixtures (*15–18*) provide the basis for interpreting the OMEGA reflectance spectra.

The detection of specific minerals from reflectance spectra involves several steps. First the atmospherically corrected data must demonstrate the presence of unique spectral features that exceed the noise or any systematic variations of the measurements. Typically, the spectral features are weak because of mixing with bright dust on Mars. In addition, there may be residual atmospheric or instrumental effects that affect the shape and strength of diagnostic absorptions and, thus, mineral interpretation. To enhance the spectral features of the important material, we find the ratio of the observed spectrum to that of a nearby dusty region with a similar atmospheric path length acquired during the same observation sequence. Spectra of bright dust exhibit no mafic mineral features, and this dust is a product of alteration processes (i.e., not fine-grained or powdered crustal material). Thus, spectral properties that are common between the measurements, including residual atmospheric effects, cancel out, leaving in the ratio the spectral properties of the material of interest. In the mineral identifications presented here, we show the original OMEGA spectra of the target terrain and nearby dusty terrain, with atmospheric correction, the ratio of target terrain to dusty terrain, and candidate spectra of minerals measured in the laboratory.

The detection of olivine is made on the basis of a broad complex of overlapping absorptions centered near 1 μm. The spectrum of the olivine-bearing surface shown in Fig. 1 exhibits a broad and strong absorption between 0.8 and 1.5 μm but is relatively featureless for wavelengths >1.5 μm. This is

[1]Geological Sciences, Brown University, Providence, RI 02912, USA. [2]Institut d'Astrophysique Spatiale, Bâtiment 121, 91405 Orsay Campus, France. [3]IDES, Bâtiment 509, 91405 Orsay Campus, France. [4]IFSI-INAF, Rome, Italy.

*To whom correspondence should be addressed. E-mail: john_mustard@brown.edu

emphasized in the ratio spectrum, where the strength of the broad mafic band relative to the dusty terrain is nearly 20%. With increasing olivine iron content, the absorptions strengthen and broaden (16). However, laboratory measurements have shown that the spectral shapes depend on both the Fe content and grain size. An increase in grain size broadens the band, shifting the position of the rise in reflectance near 1.5 μm toward longer wavelengths. Thus, a spectrum of forsteritic (low-Fe) olivine with large particles (≫100 μm) is similar to that of a fayalitic (high-Fe) olivine with fine particle size.

Figure 1 illustrates the spatial distribution of olivine-rich surfaces for an OMEGA observation. The presence of olivine is derived by a spectral index that measures the strength and breadth of the 1-μm band (19). The olivine-rich zone is on the floor of a 65-km-diameter impact crater but constrained to the east-northeast border of the crater floor. High-resolution Mars Orbiter Camera (MOC) images show that the deposit is rough, and Thermal Emission Imaging System (THEMIS) night-time observations indicate that the olivine-rich surface is the warmest in the observation, which indicates a coarse-grained, blocky, or cemented surface. This small, highly localized deposit is typical of most of the observed olivine occurrences.

Similar olivine-rich crater floor deposits are observed elsewhere on Mars, as well as isolated occurrences of olivine-rich rocks in crater rims and massifs distributed throughout

the ancient cratered terrain (Fig. 2). The surface morphology of occurrences in these regions is rugged, which indicates that the crustal rocks have been exposed by erosion, excavated from depth by impact, or subjected to other geological processes. Typically, the exposures are small and are not uniformly distributed, with more numerous occurrences in the regions richer in basalts (i.e., Syrtis Major, southeastern Valles Marineris, and Terra Cimmeria). The largest exposure identified so far is in Nili Fossae (Fig. 2), where Hoefen et al. (9) reported olivine detected with thermal IR data. Several isolated olivine-rich deposits have been observed in floors and ejecta of craters in the Northern Hemisphere between 50°N and 65°N (12).

Areas with a strong signature of LCP are identified in the OMEGA data on the basis of the position of two main absorption bands near ~0.9 μm and ~1.9 μm. As shown in Fig. 3, the OMEGA spectrum of an LCP-rich region, relative to the nearby dusty region, shows enhanced, broad 1- and 2-μm bands. The ratio spectrum emphasizes these bands and shows that their strengths are approximately 10 to 15%, relative to dusty regions. The band minima are located near 0.93 and 1.92 μm and, with reference to laboratory spectra, the surface material is dominated by LCP.

OMEGA data indicate that outcrops with high concentrations of LCP are typically located in Noachian-aged cratered terrain [see global maps in (12)]. Many are located

in crater rims, central peaks, and isolated outcrops in the cratered terrain (Fig. 3) as well as in the walls and along the floor of Valles Marineris. These outcrops rarely exceed a few square kilometers. Most of the regions occur in the southern hemisphere: The important morphogeologic aspect is the occurrence in Noachian-aged cratered terrain.

High concentrations of LCP are also identified near Syrtis Major and in Terra Meridiani. Here, the spatial extent of the deposits can reach several hundred square kilometers. The deposit near Syrtis Major is associated with the Noachian terrain located northeast of the volcanic shield (20). In Terra Meridiani, LCP is identified inside the Noachian Dissected Cratered Terrain (21).

HCP-rich areas are identified in the OMEGA data set on the basis of absorption-band minima close to 1.05 and 2.3 μm (Fig. 4). However, the observed absorptions are too wide to correspond to pure HCP (18), and thus most likely are a mixture of LCP and HCP (3, 22). We have modeled these regions using an approach based on the Modified Gaussian Model (MGM) (3, 18, 23). The spectra cannot be fit by a single-pyroxene model but are well fit by a mixture of two pyroxenes with MGM band centers appropriate for LCP and HCP. These analyses
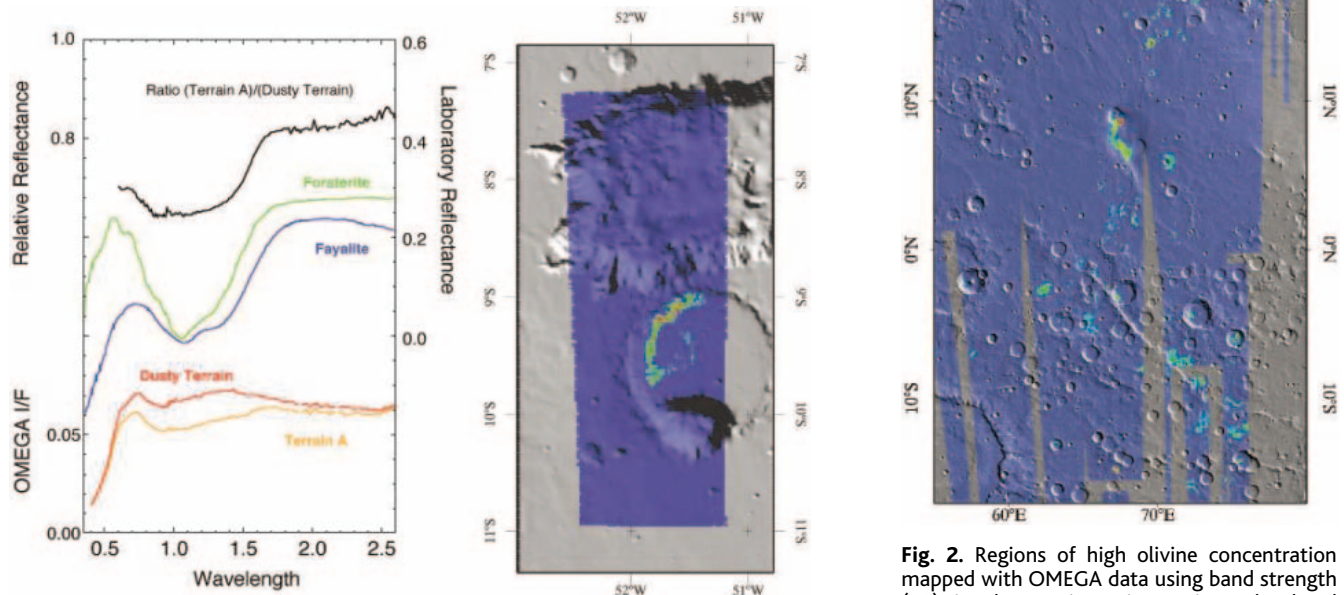
Fig. 1. Identification of Olivine. Spectral plot on left showing two OMEGA I/F atmospherically corrected spectra with target spectrum (Terrain A) and nearby spectrally neutral region (Dusty Terrain), laboratory spectra of Fe-rich (Fayalite) and Mg-rich (Forsterite) olivine, and ratio of (Terrain A)/(Dusty Terrain). On the right is a map of olivine band strength (19) from the OMEGA data showing the spatial distribution draped over Mars Orbiter Laser Altimeter (MOLA) 128 pixels per degree (ppd) shaded relief. Red indicates olivine absorption-band strengths in excess of 10%, and blue indicates no olivine detected with this parameter.
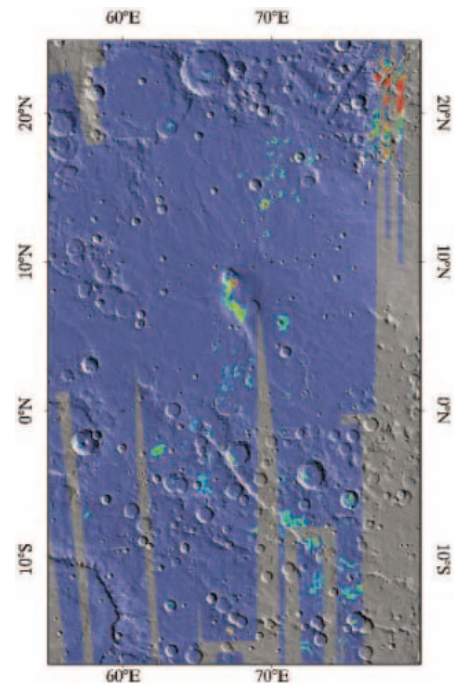
Fig. 2. Regions of high olivine concentration mapped with OMEGA data using band strength (19) in the Syrtis Major region. The band strength is draped over MOLA 128-ppd shaded relief. Red indicates olivine absorption band strengths in excess of 6% grading through green to blue, which indicates no olivine detected with this parameter. The Nili Fossae region in the northeast of this map is the largest area of olivine observed with OMEGA.

show distinct changes in the relative strength of the LCP and HCP absorptions. Laboratory-derived relationships between relative band strengths and LCP:HCP abundance (18) can be used to estimate the relative proportions of pyroxene if the pyroxenes on Mars have the same chemistries as the pyroxenes used in the laboratory studies. Although we cannot demonstrate this congruence, the MGM band centers and widths are comparable to those used in laboratory studies, and this similarity is sufficient to estimate broad differences in relative abundance. Applying the relationships of (18), the surfaces with the strongest HCP bands correspond to a mixture of 20:80 LCP:HCP, but more typically the ratio is 40:60. This relative abundance is typical of basaltic meteorites from Mars (referred to as the Shergotty, Nakala, and Chassigny, or SNC, meteorites) (24), and thus we interpret these surfaces dominated by HCP to be basaltic in composition, in agreement with the results of the Thermal Emission Spectrometer (TES) for Type I terrain (1, 2).

Many areas with total 2-μm HCP band depths greater than 10%, as determined with the MGM, extend over several hundred square kilometers [Fig. 4 and global maps shown in (12)]. Many areas enriched in HCP are associated with wind streaks, which suggests that the surface material is or was mobile (e.g., sand dunes or dark dust). Over sand dunes, a systematic enrichment in HCP relative to the surrounding areas is observed. HCP 2-μm band strengths of between 10 and 15% (determined with the MGM) are found over large regions of Mars associated with dark terrains and outcrops in the southern highlands. One notable area is the Hesperian-aged Syrtis Major volcanic shield. The spectral properties indicating an enrichment in HCP for this region are consistent with previous investigations with visible/NIR (3) and thermal IR (2). In some of the Valles Marineris canyons, layers of pyroxene-rich material with this basaltic signature are identified in one of the upper layers of the southwestern wall of Juventae Chasma and in the wall of Coprates Chasma.

The distribution of mafic minerals in the data acquired during the first 6 months of operations provides insight into the crustal composition of Mars. Olivine-rich outcrops and deposits are found in a wide range of geologic environments. They are observed in outcrops of ancient crust across a broad swath of Mars, occupying isolated massifs, crater rims, large exposures of ancient crust, and the isolated outcrops in the walls of Valles Marineris. These deposits may be due to ultramafic lava flows erupted during the formation of the early crust or may be the result of fractional crystallization and segregation in magma chambers. Olivine is abundant in the central part of the Hesperian-aged Syrtis
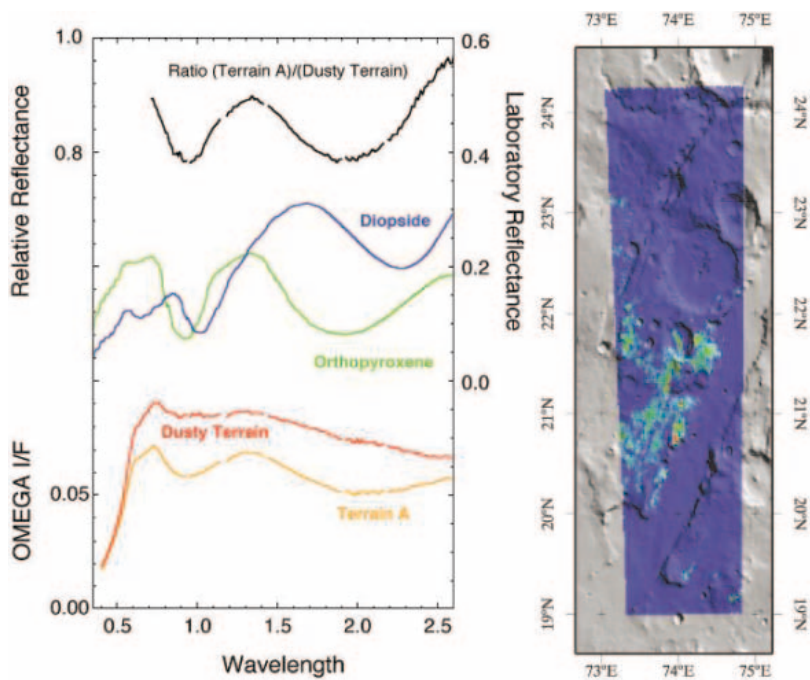
Fig. 3. Identification of LCP. Spectral plot on left showing two OMEGA I/F atmospherically corrected spectra with (a) target spectrum (Terrain A), and (b) nearby spectrally neutral region (Dusty Terrain), laboratory spectra of HCP (Diopside) and LCP (Enstatite), and ratio of (a)/(b). Next to the spectral plot is a map from the OMEGA data showing the spatial distribution draped over MOLA 128-ppd shaded relief. Red indicates LCP absorption-band strengths >10%; blue indicates absorption strengths ≤4%.
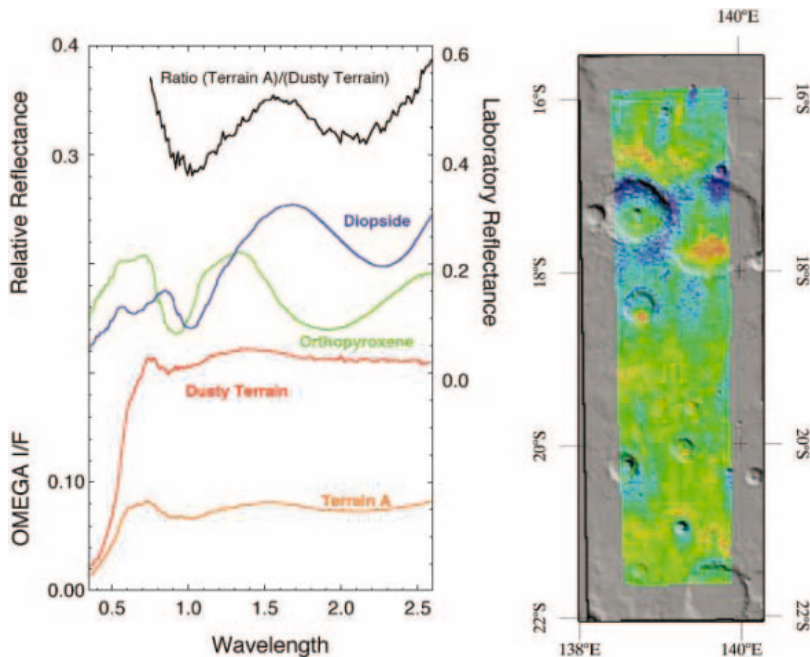
Fig. 4. Identification of HCP. Spectral plot on left showing two OMEGA I/F atmospherically corrected spectra with (a) target spectrum (Terrain A), and (b) nearby spectrally neutral region (Dusty Terrain), laboratory spectra of HCP (Diopside) and LCP (orthopyroxene), and ratio of (a)/(b). Next to the spectral plot is a map from the OMEGA data showing the spatial distribution draped over MOLA 128-ppd shaded relief. Red indicates HCP absorption-band strengths >10%; blue indicates strengths ≤4%.

Major volcanic complex (Fig. 2). These occurrences are directly associated with volcanic activity and represent mafic, ultramafic, to perhaps picritic lavas, similar to those observed in the Gusev landing site (25). The widespread occurrence of olivine-rich floors of
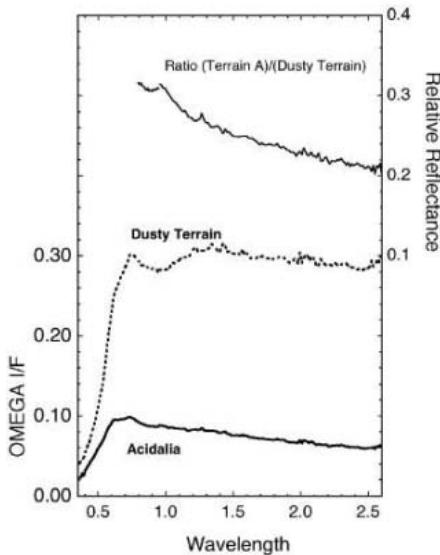
Fig. 5. OMEGA spectra from the Acidalia region of the northern lowlands of Mars. Two OMEGA I/F atmospherically corrected spectra are shown [Acidalia is from a region identified as Type II (*2*) and a nearby spectrally neutral Dusty Terrain], as well as the ratio of Acidalia/Dusty Terrain. Compared with the spectra shown in Figs. 1, 3, and 4, the Acidalia region exhibits no diagnostic spectral features.

impact craters is interesting. If they are volcanic, there are no obvious local sources for the large number of these deposits. They may instead result from impact melts.

High concentrations of LCP are found only in the ancient cratered terrain [global maps in (*12*)]. Like olivine occurrences, they are observed in central peaks, crater rims, isolated massifs, and the walls of Valles Marineris, although there is no apparent spatial correlation with olivine. LCP does occur in association with HCP in the more recent volcanic deposits in Syrtis Major and elsewhere. However, the important characteristic of the outcrops enriched in LCP is the exclusive occurrence in ancient cratered terrain. The only orthopyroxenite among the SNC meteorites is also the oldest SNC (~4.5 billion years old) and was likely formed by fractional crystallization of a basaltic parent melt (*26*). The olivine and LCP outcrops in the ancient cratered terrains are likely remnants of the ear-ly crustal formation representing cumulates created in large magma chambers or the result of melts extracted from a depleted martian mantle (*27*).

Low-albedo regions of the southern highlands are enriched in HCP relative to LCP and sometimes contain olivine [see global maps in (*12*)]. These are interpreted to be mafic volcanic rocks, likely basaltic to picritic (when olivine-rich). A two-pyroxene composition with and without olivine is typical of the basaltic SNC meteorites (*24*). This mineralogy is best expressed in Syrtis Major, which is the type locality of the basaltic compositions identified with TES (*2*). However, this spectral type is found widely in the uplands of the southern hemisphere wherever dark materials are exposed, including outcrops in crater rims in the ancient cratered terrain.

We can place these petrologic observations into a stratigraphic context representing the crust of Mars. Olivine is found from outcrops in ancient terrains to high concentrations in Hesperian-aged volcanics and crater floors. Thus, magma compositions capable of crystallizing olivine have occurred throughout much of Mars' history. High concentrations of LCP are found only in the ancient cratered terrains, and thus this rock type appears to have been formed only in the Noachian. HCP is found throughout the stratigraphic column, typically in combination with LCP and olivine, in more recent Hesperian-aged volcanic flows. These volcanic compositions indicate that the source regions were strongly depleted in aluminum and calcium.

OMEGA has sampled broad areas of the northern lowlands, the type locality for Type II materials (*2*), and also where Mars Pathfinder landed. The spectral properties of these regions (Fig. 5) are best characterized as low albedo with a strong blue slope (decreasing reflectance as a function of wavelength) from 0.9 to 2.6 μm. A weak 2-μm absorption band is observed in some areas north of Arabia, but there is no well defined or apparent 1-μm band. We have searched carefully for evidence of molecular vibration absorptions due to H_2O and/or OH^-, which might indicate the presence of alteration products and phyllosilicates. None of the materials in the areas identified as high concentrations of Type II (*2*) exhibit any absorptions due to molecular vibrations. The lack of any evidence of phyllosilicates or hydrated minerals argues against major alteration of basalt by water. The lack of distinct mafic mineral bands indicates that minerals expected in andesite are not expressed in these data. However, alteration rinds or coatings as observed at the Gusev landing site (*25*) would mask such signatures.

References and Notes
1. H. Y. McSween *et al.*, *J. Geophys. Res.* **108**, 10.1029/2003JE002175 (2003).
2. J. L. Bandfield *et al.*, *Science* **287**, 1626 (2000).
3. J. F. Mustard *et al.*, *J. Geophys. Res.* **102**, 25605 (1997).
4. V. E. Hamilton *et al.*, *J. Geophys. Res.* **106**, 14733 (2001).
5. M. B. Wyatt, H. Y. McSween, *Nature* **417**, 263 (2002).
6. R. V. Morris *et al.*, Sixth International Conference on Mars, Pasadena, CA, 20 to 25 July 2003, LPI Contribution 3211.
7. M. E. Minitti *et al.*, *J. Geophys. Res.* **107**, E5 (2002); 10.1029/2001JE001518.
8. M. D. Kraft, J. R. Michalski, T. G. Sharp, *Geophys. Res. Lett.* **30**, 2288 (2003); 10.1029/2003GL018848.
9. T. M. Hoefen *et al.*, *Science* **302**, 627 (2003).
10. V. E. Hamilton *et al.*, *Meteorit. Planet. Sci.* **38**, 871 (2003).
11. J.-P. Bibring *et al.*, *ESA SP* **1240**, 37 (2004).
12. J.-P. Bibring *et al.*, *Science* **307**, 1576 (2005); 10.1126/science.1108806.
13. J.-P. Bibring *et al.*, *Nature* **341**, 591 (1989).
14. R. G. Burns, *Mineralogic Applications of Crystal Field Theory* (Cambridge Univ. Press, Cambridge, 1970).
15. J. B. Adams, *J. Geophys. Res.* **79**, 4829 (1974).
16. T. V. V. King, I. Ridley, *J. Geophys. Res.* **92**, 11457 (1987).
17. E. A. Cloutis, M. J. Gaffey, *J. Geophys. Res.* **96**, 22809 (1991).
18. J. M. Sunshine *et al.*, *J. Geophys. Res.* **98**, 9075 (1993).
19. The spectral index to map olivine is calculated from reduced, atmospherically corrected OMEGA data with the following formula: Index = [R1700/(0.1*R1020 + 0.1*R1210 + 0.4*R1330 + 0.4*R1470)], where R1700 is the reflectance at 1700 nanometers, R1020 is the reflectance at 1020 nanometers, and so on.
20. R. Greeley, J. E. Guest, Geological Map of the Eastern Equatorial Region of Mars, U.S. Geol. Surv. Misc. Inv. Series Map I-1802-B (1987).
21. R. E. Arvidson *et al.*, *J. Geophys. Res.* **108**, 8073 (2003).
22. J. F. Mustard, J. M. Sunshine, *Science* **267**, 1623 (1995).
23. J. M. Sunshine, C. M. Pieters, *J. Geophys. Res.* **95**, 6955 (1990).
24. H. Y. McSween, *Meteorit. Planet. Sci.* **37**, 7 (2002).
25. H. Y. McSween *et al.*, *Science* **305**, 842 (2004).
26. D. W. Mittlefehldt, *Meteoritics* **29**, 900 (1994).
27. P. C. Hess, Workshop on Unmixing the SNCs (Lunar and Planetary Institute, Houston, TX, 11 to 12 October 2002), LPI Contribution 6029.
28. J.F.M. acknowledges NASA support for this work. We thank the personnel involved in acquisition and reduction of Mars Express OMEGA data for their efforts.

BREVIA

Isolation of an Algal Morphogenesis Inducer from a Marine Bacterium

Yoshihide Matsuo,[1]* Hiroshi Imagawa,[2] Mugio Nishizawa,[2] Yoshikazu Shizuri[1]

Ulva (sea lettuce) and *Enteromorpha* (gut weed) are cosmopolitan, intertidal marine algae that display a foliaceous morphology in their natural environment. However, these green macroalgae lose their natural morphology during short-term cultivation under aseptic conditions and during long-term cultivation in nutrient-added seawater (*1–7*). This observation has led to the belief that undefined morphogenetic factors are present in the ocean that are indispensable for the foliaceous morphology of macroalgae (*2*).

The alga *Monostroma oxyspermum* (Monostromataceae, Ulvales) has a leafy morphology in its natural state and looks like thin *Ulva*. However, under aseptic conditions, it grows as loose aggregates of single cells (*7*). We have shown previously that these cells can redifferentiate upon reinfection with marine bacteria from the Cytophaga-Flavobacterium-Bacteroides (CFB) group that were isolated from the green algal surface (*8*). We have also demonstrated that bacterial extracellular metabolites can induce not only the normal morphogenesis of *M. oxyspermum* but also the normal germination of germfree spores of *Ulva pertusa* and *Enteromorpha intestinalis* (*8*).

We report here the isolation of a highly potent differentiation inducer, termed thal-lusin (Fig. 1A), from the epiphytic marine bacterium strain YM2-23 (deposition no. MBIC 04683), belonging to the CFB group that was isolated from *Monostroma* sp. (*8*). The structure of thallusin was determined by spectral and single-crystal x-ray structure determination of a thallusin derivative, Me1H1W4 (fig. S1) (*9*).

Pure thallusin strongly induced the differentiation of *M. oxyspermum* (Fig. 1, B and C). The minimum effective concentration (*8*) in a cell differentiation assay with *M. oxyspermum* was between 1 fg ml^{-1} and 1 ag ml^{-1}, depending on the lighting conditions, the assay duration (usually 4 days), and the initial number of *M. oxyspermum* cells (usually 20 cells per ml). When thallusin was added at a concentration of 100 μg l^{-1}, no inhibitory activity was apparent in the cell differentiation or growth of *M. oxyspermum*. The morphogenesis of *M. oxyspermum* into the thallus with thallusin addition and the gradual disintegration of the thallus into single cells with thallusin deprivation (Fig. 1D) suggest that thallusin must be continuously supplied by an epipythic marine bacterium to maintain algal body.

Thallusin was also able to induce the germination of such other green macroalgae as *U. pertusa* (Fig. 1, E and F) and *E. in-*testinalis (*10*), after they had been cultured under aseptic conditions. Previous studies have also reported that these green algae did not develop as normal foliate gametophytes under aseptic conditions (*1–5*). Because thallusin was effective toward *U. pertusa* and *E. intestinalis*, in addition to *M. oxyspermum*, thallusin-dependent morphogenesis may be a common feature in Monostromataceae and Ulvaceae (*8*).

The mode of action of the algal response to thallusin is unknown, but a bacterial supply of thallusin functions in the development of algal communities. Other differentiation inducers may be found in nature. The role of thallusin shown here exemplifies the fundamental symbiotic chemical communication between macroalgae and epiphytic bacteria in the marine environment.

References and Notes

1. L. Provasoli, *Biol. Bull.* **144**, 375 (1958).
2. E. R. Bonneau, *J. Phycol.* **13**, 133 (1977).
3. L. Fries, *Bot. Mar.* **18**, 251 (1975).
4. L. Provasoli, *Proc. Int. Seaweed Symp.* **4**, 1 (1963).
5. K. Nakanishi *et al.*, *J. Phycol.* **32**, 479 (1996).
6. L. Provasoli, I. J. Pintner, *J. Phycol.* **16**, 196 (1980).
7. M. Tatewaki, L. Provasoli, I. J. Pintner, *J. Phycol.* **19**, 409 (1983).
8. Y. Matsuo *et al.*, *Environ. Microbiol.* **5**, 25 (2003).
9. Materials and methods are available as supporting material on *Science* Online.
10. Y. Matsuo, H. Imagawa, M. Nishizawa, Y. Shizuri, unpublished data.
11. We thank M. Suzuki and M. Tatewaki for their valuable suggestions and for providing the *M. oxyspermum* strain, K. Nakanishi for critical discussions of this manuscript, and the two reviewers for helpful comments. Supported in part by the New Energy and Industrial Technology Development Organization (NEDO). The 16S ribosomal DNA and DNA gyrase B gene (*gyrB*) sequences of strain YM2-23 (MBIC 04683) have been deposited at the DNA Data Bank of Japan/GenBank/European Molecular Biology Laboratory data banks under respective accession nos. AB073558 and AB073149. The crystal structure of Me1H1W4 has been deposited at the Cambridge Crystallographic Data Centre with deposition no. CCDC234022.

[1]Marine Biotechnology Institute Co. Ltd., 3-75-1 Heita, Kamaishi-shi, Iwate 026-0001, Japan. [2]Faculty of Pharmaceutical Sciences, Tokushima Bunri University, Yamashiro-cho, Tokushima 770-8514, Japan.

*To whom correspondence should be addressed. E-mail: yoshihide.matsuo@mbio.jp

Fig. 1. (**A**) The structure of thallusin. Only the relative configuration was determined. (**B**) Morphology of axenic *M. oxyspermum*, after 7 days of cultivation. Inset: A loose culture mass consisting of single cells. (**C**) The effect of thallusin on *M. oxyspermum* (an initial concentration of 1 ng ml^{-1}). Inset: A typical monostromatic gametophyte. (**D**) Dedifferentiation given a lack of thallusin (an initial concentration of 1 pg ml^{-1}). Inset: Protruding cells from disintegrating thalli. (**E**) Small callus-like morphology of *U. pertusa*, after 2 months of cultivation under aseptic conditions. Inset: Colorless protrusions from the lateral cell walls. (**F**) The effect of thallusin (1 ng ml^{-1}) on *U. pertusa*. Inset: A foliaceous distromatic blade. Scale bars, (B) to (D), 100 μm; (E) and (F), 4 mm; insets, 20 μm.
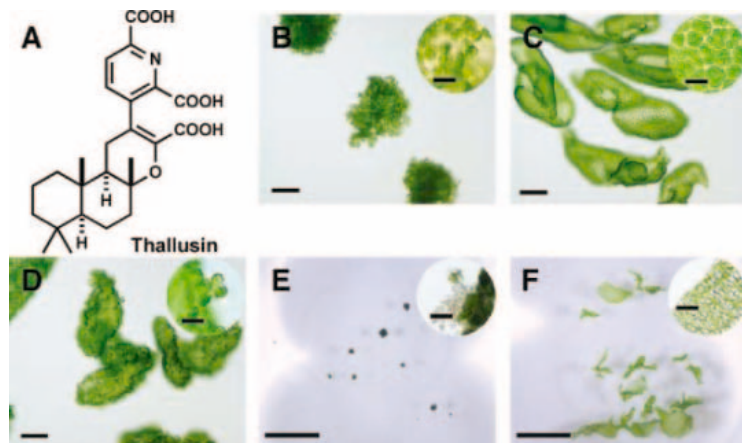
Chaperone Activity of Protein O-Fucosyltransferase 1 Promotes Notch Receptor Folding

Tetsuya Okajima, Aiguo Xu, Liang Lei, Kenneth D. Irvine*

Notch proteins are receptors for a conserved signaling pathway that affects numerous cell fate decisions. We found that in *Drosophila*, Protein O-fucosyltransferase 1 (OFUT1), an enzyme that glycosylates epidermal growth factor–like domains of Notch, also has a distinct Notch chaperone activity. OFUT1 is an endoplasmic reticulum protein, and its localization was essential for function in vivo. OFUT1 could bind to Notch, was required for the trafficking of wild-type Notch out of the endoplasmic reticulum, and could partially rescue defects in secretion and ligand binding associated with Notch point mutations. This ability of OFUT1 to facilitate folding of Notch did not require its fucosyltransferase activity. Thus, a glycosyltransferase can bind its substrate in the endoplasmic reticulum to facilitate normal folding.

The extracellular domain of Notch proteins consists largely of tandemly repeated epidermal growth factor–like (EGF) domains. Notch folding must be tightly controlled, because disruption of a single EGF domain can dominantly perturb Notch, as in the human congenital syndrome CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy). Notch receptors are modified within their EGF domains by the addition of fucose to Ser or Thr residues (O-fucose) (*1*, *2*). *Drosophila* Notch contains 36 EGF domains, 23 of which can potentially be O-fucosylated. O-fucosylation is catalyzed by Protein O-fucosyltransferase 1 (O-FucT-1), which is encoded by *Ofut1* in *Drosophila* and by *Pofut1* in mammals (*3*). *Ofut1* and *Pofut1* are essential for Notch signaling in vivo (*4–7*).

Notch accumulates in the ER in OFUT1-depleted cells. We expressed an *Ofut1* RNA interference (RNAi) construct, *iOfut1*, in a stripe of cells in *Drosophila* wing discs and confirmed that OFUT1 protein levels were decreased (Fig. 1A). This resulted in elevated Notch protein staining, as also seen in clones of *Ofut1* mutant cells (*4*) (Fig. 1B) (fig. S1). *Notch* transcription was not affected by *Ofut1* RNAi in wing discs (Fig. 1C). Activation of Notch by its ligands induces Notch processing (*8*), which suggests that the accumulation of Notch might derive from its inability to interact with ligands. However, clones of cells mutant for genes required for ligand-dependent cleavages of Notch (*9*) or doubly mutant for the two Notch ligands *Delta* and *Serrate* (Fig. 1D) did not result in elevated Notch protein. Thus, OFUT1 has an influence on Notch that is distinct from any requirement for Notch-ligand binding (*5*, *7*).

In wild-type wing cells, Notch is localized near the adherens junctions, similar to the apical protein Crumbs (*10*, *11*) (Fig. 1E). In OFUT1-depleted cells, Notch accumulated throughout the cytoplasm both basally and apically (Fig. 1, B and E). By comparison with Crumbs, Notch levels at the apical membrane were decreased relative to wild-type cells (Fig. 1E). Crumbs contains 30 extracellular EGF domains, many of which could potentially be O-fucosylated (*2*, *11*), and the localization of the other known substrates of OFUT1, Delta and Serrate (*12*), was not changed in OFUT1-depleted cells (fig. S1). Thus, OFUT1 is specifically required for cell surface localization of Notch.

To identify where Notch trafficking becomes impaired, we compared the localization of Notch to markers of subcellular compartments. Notch staining in cells lacking

Howard Hughes Medical Institute, Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, USA.

*To whom correspondence should be addressed. E-mail: irvine@waksman.rutgers.edu

Fig. 1. OFUT1 depletion results in accumulation of Notch in the ER. Wing discs are shown with dorsal up and anterior to the left. Panels (A) to (C) and (E) and (F) show expression of *iOfut1* under *ptc-Gal4* control (*UAS-iOfut1 ptc-Gal4*). In this and later figures, panels marked with primes show separate channels of the same disc. Additional controls are in fig. S1. (A) Disc stained with antiserum to OFUT1 (brown) reveals depletion effected by *iOfut1*; inset shows magnification of the boxed area. (B) Disc stained with anti-Notch (magenta) reveals elevated levels in OFUT1-depleted cells. Presence of a *UAS-GFP* transgene allows GFP (green fluorescent protein) to mark *ptc-Gal4*–expressing cells. Vertical section is shown at bottom. The boxed areas labeled "E,F" and "G" show approximate locations of images in (E) and (F) and in fig. S1, e to h (G), respectively. (C) In situ hybridization to *Notch* mRNA. (D) *Delta Serrate* mutant clones (green, white arrow) do not exhibit elevated Notch staining. Vertical section is shown at bottom. (E) Localization of Notch, but not Crumbs, is altered in OFUT1-depleted cells; note the paucity of white staining on the left side, reflective of the absence of Notch at the apical membrane. (F) Notch overlaps with an ER marker, anti-KDEL, in OFUT1-depleted cells. The dashed line in (E) and (F) marks the edge of the *ptc* expression stripe. (E) presents a more apical section than (F); mislocalization of Notch is most evident in (E), whereas elevated levels are most evident in (F).

OFUT1 overlapped markers of the endoplasmic reticulum (ER) (anti-KDEL and Boca) (Fig. 1F) (fig. S1), but did not overlap markers for Golgi (anti-Golgi), early endosomes (Hrs, GFP:rab5), or late endosomes (GFP:rab7) (fig. S1). Cultured *Drosophila* S2 cells express OFUT1 but lack a functional *Notch* gene. A fraction of transfected Notch in S2 cells appeared on the cell surface, whereas the rest remained in the ER (Fig. 2A) (fig. S2). When OFUT1 was depleted from S2 cells by RNAi, cell surface staining of Notch was eliminated and only ER staining remained (Fig. 2B). We also examined a ~120-kD Notch proteolytic fragment, which is thought to be generated by processing steps subsequent to secretion out of the ER (*13*). In OFUT1-depleted cells, this processed form of Notch was undetectable (Fig. 2F). We also assayed the secretion into the medium of a Notch–alkaline phosphatase (N:AP) fusion protein consisting of the 36 EGF domains of Notch fused to AP. A strong decrease in N:AP secretion was observed in OFUT1-depleted cells, whereas secretion of the analogous Delta and Serrate fusion proteins DL:AP and SER:AP was much less affected (Fig. 2G). The deficit in Notch secretion reflects a specific requirement for an O-FucT-1, because it can be reversed by

cotransfection with mouse *Pofut1* (Fig. 2). Thus, OFUT1 is required for normal trafficking of Notch out of the ER.

OFUT1 is an ER protein. The influence of OFUT1 on Notch in the ER was unexpected, because transporter activity for guanosine diphosphate (GDP)–fucose, the donor substrate for fucosyltransferases, has been detected only on Golgi membranes (*14*), and all other fucosyltransferases examined to date are Golgi proteins (*15*). However, mammalian O-FucT-1 is modified with high mannose–type *N*-glycans, which could be consistent with localization in the ER (*3*). Indeed, in S2 cells, OFUT1 colocalized with ER markers (GFP:KDEL and anti-KDEL) (Fig. 3A) and did not overlap with Golgi markers (Fig. 3B). Similarly, in the wing disc, OFUT1 partially colocalized with an ER marker (anti-KDEL, Fig. 3C).

Soluble ER proteins are retained by a retrograde transport system that recognizes the C-terminal tetrapeptide KDEL and related variants. The presence of KDEL-like sequences is conserved in OFUT1 orthologs (Fig. 3D), which suggests that OFUT1 is a soluble ER protein. When OFUT1 was overexpressed, a small amount was secreted, consistent with observations that the KDEL receptor can be saturated (*16*); also, the mobility of OFUT1

on SDS–polyacrylamide gel electrophoresis (PAGE) gels in the culture media was indistinguishable from that in cell lysates, as expected if the N terminus is a signal peptide (fig. S3). Moreover, the N terminus of OFUT1 could be substituted with a signal peptide from another protein, Bip, without affecting its localization. Available antisera are directed against the C terminus of OFUT1 (*5*), but we were able to insert an S-epitope tag into Bip:OFUT1 without affecting its localization or activity; hence, we used this chimera to examine the requirement for the C-terminal HEEL (*17*) sequence. Deletion of these four amino acids substantially increased secretion of OFUT1 into the medium (Fig. 3, E and F).

A mutation in *Ofut1*, *Ofut1^{SH}*, causes a severe loss-of-function phenotype in flies and is associated with replacement of the seven C-terminal amino acids of OFUT1 by four other amino acids (*5*). Because the C terminus includes an ER retention signal, this mutation might derive from improper localization rather than loss of enzymatic activity. To test this idea, we created a variant of Bip:OFUT1 carrying the C-terminal *Ofut1^{SH}* mutation. This protein was secreted into the medium as efficiently as was Bip:OFUT1 lacking the HEEL sequence (Fig. 3E). Moreover, it displayed substantial enzymatic activity (Fig. 3F). Thus, *Ofut1^{SH}* is a mislocalization mutant, and retention in the ER is required for normal OFUT1 function in vivo.

Fucosyltransferase activity of OFUT1 is not required for Notch secretion. Because the ER lacks detectable GDP-fucose transporter activity, we thought that OFUT1 might facilitate Notch folding independently of its fucosyltransferase activity. A comparative study of sequences from α1,2-fucosyltransferases and α1,6-fucosyltransferases identified a GXHXR(R/H) motif that is implicated in donor substrate binding (*18*, *19*); O-fucosyltransferases include a related motif (*20*). The first R in this motif is invariant among all three fucosyltransferase families (Fig. 4A) and is essential for the activity of α1,6- and α1,2-fucosyltransferases in vitro and in vivo (*18*, *19*, *21*). We mutated this R in OFUT1, making OFUT1^{R245K} and OFUT1^{R245A}. Secreted versions of these mutants, including a C-terminal V5-hexahistidine tag, were expressed in S2 cells. The mutant chimeras were secreted as well as was the wild-type chimera, which suggests that they folded normally (fig. S3), but lacked fucosyltransferase activity (Fig. 4E) (fig. S3).

We then investigated the ability of OFUT1^{R245A} to promote Notch secretion. In developing wing discs, we created clones of cells that were simultaneously mutant for a null allele of the endogenous *Ofut1* gene and expressed OFUT1^{R245A} under the control of a heterologous promoter. Expression of this enzymatically inactive form of OFUT1 restored

Fig. 2. OFUT1 is essential for secretion of Notch in S2 cells. (**A** to **E**) S2 cells transfected to express Notch, stained before detergent treatment for cell surface Notch (ECN, red) and after detergent for intracellular Notch (ICN, green). In (B) to (E), cells were also treated with double-stranded RNA (dsRNA) corresponding to *Ofut1*; in (C) to (E), cells were also cotransfected with Pofut1 (C), Pofut1^{R245A} (D), or Boca (E). (**F**) Immunoblot of S2 cell lysates probed with antibodies to the Notch intracellular domain and to OFUT1 and tubulin (as controls). Where indicated, lysates are from cells transfected with Notch, Boca, Pofut1, or Pofut1^{R245A} and treated with *Ofut1* dsRNA. Upper arrow points to full-length Notch; lower arrow points to a 120-kD fragment whose appearance requires OFUT1. (**G**) Secretion of AP-tagged Notch, Delta, or Serrate extracellular domains. Relative AP activity secreted into conditioned media is shown from transfected S2 cells, cells treated with dsRNA corresponding to *Ofut1*, and, where indicated, cells cotransfected with Boca, Pofut1, or Pofut1^{R245A}. AP activity was normalized for transfection efficiency by luciferase activity in lysates (AP/luc).
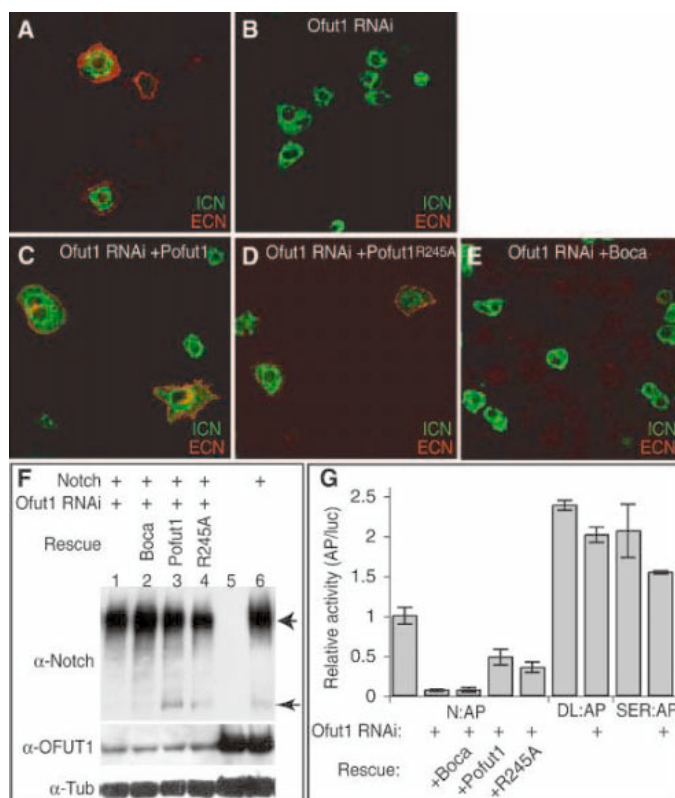
Notch localization to the cell surface and partially suppressed intracellular staining of Notch (Fig. 4, B and C). To conduct rescue experiments in S2 cells, we engineered the R245A mutation into murine Pofut1 (Pofut1[R245A]) (fig. S3). Expression of Pofut1[R245A] in OFUT1-depleted S2 cells partially restored the trafficking of Notch to the cell surface, as assayed by immunostaining, production of the 120-kD cleavage product, and secretion of N:AP (Fig. 2). A control ER protein, the low-density lipoprotein (LDL) receptor chaperone Boca, could not restore Notch secretion (Fig. 2). Thus, OFUT1 can facilitate the secretion of Notch independently of its fucosyltransferase activity.

As an independent test of the lack of requirement for *O*-fucose in Notch secretion, we created and analyzed mutations in the predicted *Drosophila* GDP-mannose 4,6-dehydratase (*Gmd*) (Fig. 4F), which is required for GDP-fucose biosynthesis and hence for all forms of fucosylation in *Drosophila*. A null mutant, *Gmd1*, could survive to third instar. These animals exhibited severely decreased growth of the wing disc and lacked expression of Notch target genes at the dorsal-ventral boundary [e.g., Wingless (WG)] (Fig. 4I). These phenotypes are consistent with the expected consequences of absence of *O*-fucose and were similar to those observed in *Ofut1SH* (Fig. 4H). However, in *Ofut1SH* mutants Notch accumulated in the ER (Fig. 4D), but in *Gmd1* mutants, Notch did not accumulate in the ER and could be detected on the cell surface (Fig. 4J).

OFUT1 binds Notch. The failure of Notch to be secreted from the ER in the absence of OFUT1 could reflect a chaperone activity analogous to the action of Boca, which facilitates the folding of LDL receptors by transiently associating with a specific protein domain (*16, 22, 23*). In the absence of OFUT1, misfolded Notch would then be retained in the ER by quality control mechanisms (*24, 25*). This hypothesis predicts that it should be possible to detect an association between Notch and OFUT1. Mammalian O-FucT-1 has affinity for EGF domains, as it was partially purified by EGF domain affinity chromatography (*3*). To detect binding between OFUT1 and *Drosophila* Notch, we coexpressed OFUT1:V5:His and N:AP in S2 cells and then used coimmunoprecipitation to assay the association between Notch and OFUT1. Specific binding of OFUT1 or OFUT1[R245A] to N:AP was reproducibly detected (Fig. 5A), as was weak binding to Delta (fig. S4).

Chaperone activity of OFUT1 can promote the folding of wild-type and mutant Notch. To further evaluate the hypothesis that OFUT1 is a Notch chaperone, we assayed its influence on the secretion and ligand-binding activity of Notch.

These experiments were conducted with N:AP so that amounts of Notch could be quantified (*5, 26*). OFUT1[R245A] enhanced both N:AP secretion and ligand binding (Fig. 5, B and C) (fig. S4), which suggests that at endogenous levels of OFUT1, a fraction of N:AP is misfolded. We also examined the influence of OFUT1 on mutant forms of Notch. N:AP-EGF23-32f is a form of N:AP that contains single base changes in each of the EGF domains 23 through 32, such that the site of *O*-fucose attachment is changed from the normal Ser or Thr to Ala or Val. This mutant was secreted at less than one-tenth the rate of wild-type N:AP, and it bound Delta-expressing cells less than half as well as did wild-type N:AP (Fig. 5). OFUT1[R245A] promoted the secretion and ligand binding of N:AP-EGF23-32f (Fig. 5).

CADASIL is a disease characterized by strokes, dementia, and early death; it is caused by point mutations in Notch3 (*27*) that either add or eliminate a Cys in one of the EGF domains. It is not yet clear whether the etiology of CADASIL results from alteration of normal Notch3 function or from the accumulation of protein aggregates (*28, 29*). In any case, the molecular nature of CADASIL mutations implies that they influence protein folding. By sequence alignment, we modeled two CADASIL mutations, R141C and C542Y, onto *Drosophila* Notch, and created the equivalent mutations, S200C and C599Y. Secretion and ligand binding of N:AP[S200C] were not significantly different from that of wild-type N:AP. However, both secretion and ligand binding of N:AP[C599Y] were impaired, and coexpression with OFUT1[R245A] enhanced N:AP[C599Y] secretion and ligand binding (Fig. 5). Thus, OFUT1[R245A] promoted the secretion and activity of both wild-type and mutant forms of Notch, supporting
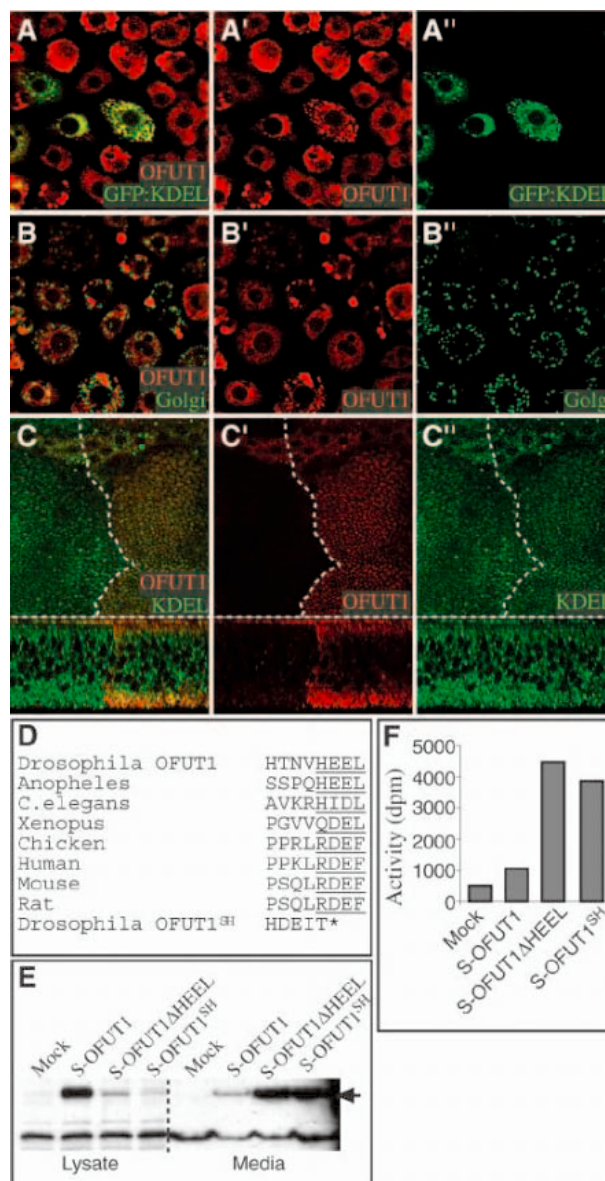
Fig. 3. OFUT1 is an ER protein. (**A**) S2 cells transfected with an ER marker (GFP:KDEL, green) and stained for OFUT1 (red). (**B**) S2 cells stained for Golgi (anti-Golgi, green) and OFUT1. (**C**) Horizontal (top) and vertical (bottom) sections of a *ptc-Gal4 UAS-iOfut1* wing disc stained for ER (anti-KDEL, green) and OFUT1. OFUT1 was depleted to the left of the dashed lines by RNAi. (**D**) Alignment of C-terminal sequences of various O-FucT-1s, with proposed KDEL signal underlined. (**E**) Immunoblot with antibody to S-tag epitope, showing localization of S-tagged OFUT1 and two OFUT1 mutants: deletion of the C terminus (ΔHEEL) and recreation of the OFUT1[SH] allele. The localization of both mutants is shifted from cell lysate to culture medium. (**F**) Assays on a Factor VII substrate of OFUT1 activity in conditioned media from cells expressing wild-type or mutant forms of OFUT1.

the conclusion that OFUT1 is a Notch chaperone.

Chaperone activity and glycosyltransferases. The observations that OFUT1 binds to Notch, that OFUT1 is an ER protein, and that Notch secretion requires OFUT1 (but not its fucosyltransferase activity) together suggest a model for OFUT1 action (fig. S5). OFUT1 could associate with folded EGF domains of Notch in the ER during folding of the extracellular domain. This association could be required to prevent inappropriate inter- or intramolecular associations of EGF domains, which would otherwise interfere with further Notch folding.

OFUT1 would catalyze the transfer of fucose either during folding (if O-fucosylation can occur in the ER) or subsequently (if it happens in another compartment). In the absence of OFUT1, individual EGF domains could still fold, but the extracellular domain would not assemble correctly, resulting in misfolded and/or aggregated Notch. Implicit in this model is that the specificity of OFUT1 for folding of Notch lies in the need to block inappropriate associations of Notch EGF domains. By contrast, the folding reactions for other EGF domain–containing proteins, such as Serrate, Delta, and Crumbs, apparently do not

present similar possibilities for inappropriate association and hence do not require OFUT1.

Other glycosyltransferases with specific protein substrates have been identified, including O-FucT-2 (30), O-glucosyltransferase (31), O-mannosyltransferase (32), and C-mannosyltransferase (33). O-and C-mannosyltransferases appear to be ER glycosyltransferases (32, 34), and the subcellular distributions of O-FucT-2 and O-glucosyltransferase and have not yet been determined. Thus, ER glycosyltransfer-
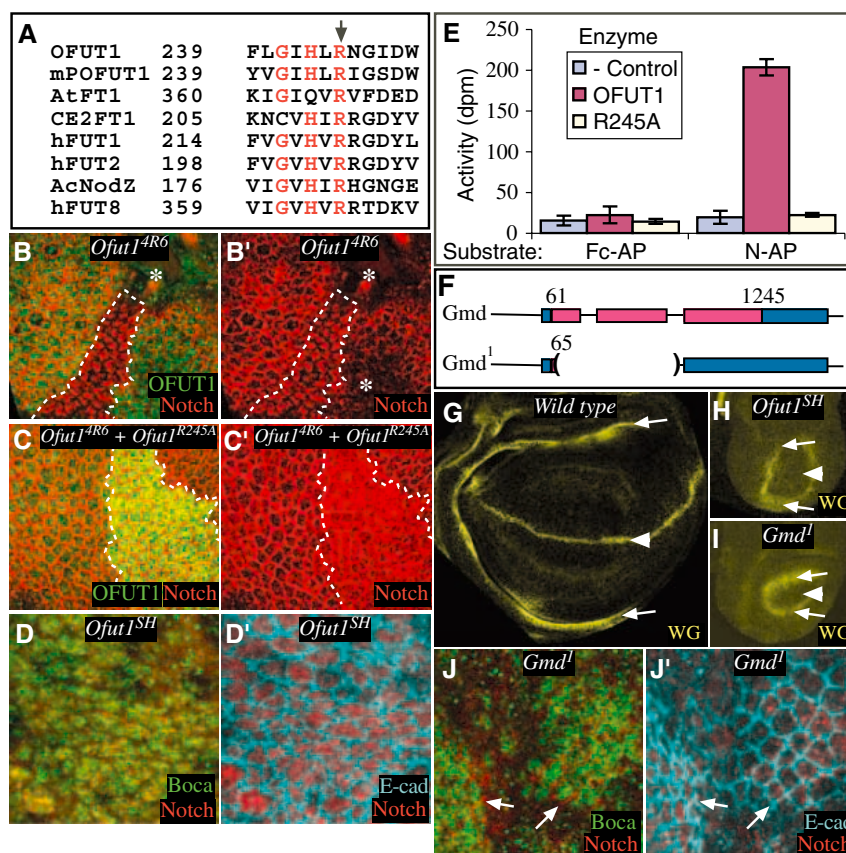
Fig. 4. Fucosyltransferase activity is not required for Notch secretion. (**A**) Alignment of α1,2-fucosyltransferases, α1,6-fucosyltransferases, and O-fucosyltransferases identifies an invariant Arg residue (arrow) in a region implicated in GDP-fucose binding (18). (**B**) Clone of cells mutant for Ofut1⁴ᴿ⁶ and marked by the absence of OFUT1 staining (green). Notch is mislocalized in mutant cells (outlined by dashes); elevated levels are only visible basally. Asterisks identify regions where Notch is out of the plane of focus. (**C**) Clone mutant for Ofut1⁴ᴿ⁶ and simultaneously expressing OFUT1ᴿ²⁴⁵ᴬ under tubulin-Gal4 control. Notch secretion to the cell surface is restored, and accumulation inside cells is reduced. Antiserum to OFUT1 recognizes OFUT1ᴿ²⁴⁵ᴬ; the clone is marked by elevated staining (green). (**D**) Notch protein localization (red) in Ofut1ˢᴴ overlaps an ER marker (Boca, green) and is complementary to an apical membrane marker (E-cadherin, cyan). (**E**) Assays on purified N:AP or control (Fc:AP) substrates of fucosyltransferase activity, using affinity-purified OFUT1:V5, OFUT1ᴿ²⁴⁵ᴬ:V5, or a mock purification (-Control) from untransfected cells as enzyme sources. (**F**) Schematic of the Gmd locus; exons are shown as bars and the open reading frame (nucleotides 61 to 1245) is in pink. In Gmd¹ DNA, a deletion begins at nucleotide 65 and ends in an intron; the predicted peptide encoded by this allele is only two amino acids. (**G** to **I**) Wing imaginal discs stained for expression of WG (yellow) from wild-type (G), Ofut1ˢᴴ (H), and Gmd¹ (I). WG at the D-V boundary (arrowhead) is regulated by Notch and is missing from Ofut1 and Gmd mutants. WG in the hinge (arrows) is regulated separately but illustrates reduced growth. (**J**) Notch protein localization in Gmd¹ is largely complementary to Boca and partially overlaps E-cadherin. Arrows highlight Notch at the apical membrane.
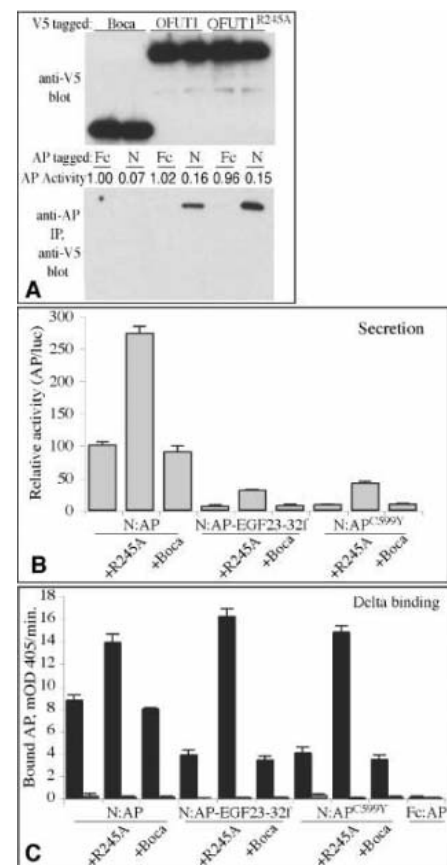
Fig. 5. Chaperone activity of OFUT1. (**A**) Coimmunoprecipitation experiment with V5-tagged OFUT1, OFUT1ᴿ²⁴⁵ᴬ, or Boca, and N:AP or Fc:AP. The indicated proteins were coexpressed in S2 cells and isolated from conditioned media. Input amounts of V5-tagged proteins are shown by immunoblot (top) and of AP-tagged proteins by AP assay. Fc:AP is expressed and secreted more robustly than N:AP, but nonetheless it does not bind OFUT1. Coimmunoprecipitation is shown by the immunoblot at bottom. (**B**) Relative secretion of N:AP and N:AP mutants in S2 cells, expressed as a normalized ratio of AP to luciferase activity when coexpressed with OFUT1ᴿ²⁴⁵ᴬ (+R245A) or Boca. (**C**) Binding of wild-type or mutant forms of N:AP (or, as a negative control, Fc:AP) to Delta-expressing (left of each pair of bars) or control S2 cells (right of each pair). AP fusions were transfected into wild-type (OFUT1-expressing) S2 cells; where indicated, cells were cotransfected with OFUT1ᴿ²⁴⁵ᴬ or Boca. Equal amounts of protein (AP activity 4000 mOD/min) were used in each case.

ases with substrate specificity might constitute a class of quality control proteins.

References and Notes

1. N. Haines, K. D. Irvine, *Nat. Rev. Mol. Cell Biol.* **4**, 786 (2003).
2. R. S. Haltiwanger, *Curr. Opin. Struct. Biol.* **12**, 593 (2002).
3. Y. Wang, M. W. Spellman, *J. Biol. Chem.* **273**, 8112 (1998).
4. T. Okajima, K. D. Irvine, *Cell* **111**, 893 (2002).
5. T. Okajima, A. Xu, K. D. Irvine, *J. Biol. Chem.* **278**, 42340 (2003).
6. S. Shi, P. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5234 (2003).
7. T. Sasamura *et al.*, *Development* **130**, 4785 (2003).
8. F. Schweisguth, *Curr. Biol.* **14**, R129 (2004).
9. Mutations in kuzbanian and presenilin were examined (T. Okajima *et al.*, data not shown).
10. R. G. Fehon, K. Johansen, I. Rebay, S. Artavanis-Tsakonas, *J. Cell Biol.* **113**, 657 (1991).
11. U. Tepass, C. Theres, E. Knust, *Cell* **61**, 787 (1990).
12. V. M. Panin *et al.*, *J. Biol. Chem.* **277**, 29945 (2002).
13. C. M. Blaumueller, H. Qi, P. Zagouras, S. Artavanis-Tsakonas, *Cell* **90**, 281 (1997).
14. L. Puglielli, C. B. Hirschberg, *J. Biol. Chem.* **274**, 35596 (1999).
15. D. J. Becker, J. B. Lowe, *Glycobiology* **13**, 41R (2003).
16. J. C. Hsieh *et al.*, *Cell* **112**, 355 (2003).
17. Single-letter abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; X, any amino acid.
18. T. Takahashi *et al.*, *Glycobiology* **10**, 503 (2000).
19. V. Chazalet, K. Uehara, R. A. Geremia, C. Breton, *J. Bacteriol.* **183**, 7067 (2001).
20. I. Martinez-Duncker, R. Mollicone, J. J. Candelier, C. Breton, R. Oriol, *Glycobiology* **13**, 1C (2003).
21. D. P. Chen *et al.*, *Ann. Clin. Lab. Sci.* **34**, 314 (2004).
22. J. Culi, R. S. Mann, *Cell* **112**, 343 (2003).
23. J. Culi, T. A. Springer, R. S. Mann, *EMBO J.* **23**, 1372 (2004).
24. B. Kleizen, I. Braakman, *Curr. Opin. Cell Biol.* **16**, 343 (2004).
25. E. S. Trombetta, A. J. Parodi, *Annu. Rev. Cell Dev. Biol.* **19**, 649 (2003).
26. K. Bruckner, L. Perez, H. Clausen, S. Cohen, *Nature* **406**, 411 (2000).
27. A. Joutel, E. Tournier-Lasserve, *Semin. Cell Dev. Biol.* **9**, 619 (1998).
28. H. Karlstrom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17119 (2002).
29. A. Joutel, M. Monet, V. Domenga, F. Riant, E. Tournier-Lasserve, *Am. J. Hum. Genet.* **74**, 338 (2004).
30. R. Haltiwanger, personal communication.
31. L. Shao, Y. Luo, D. J. Moloney, R. Haltiwanger, *Glycobiology* **12**, 763 (2002).
32. H. Manya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 500 (2004).
33. J. Hofsteenge *et al.*, *J. Biol. Chem.* **276**, 6485 (2001).
34. J. Perez-Vilar, S. H. Randell, R. C. Boucher, *Glycobiology* **14**, 325 (2004).
35. Molecular interaction data have been deposited in the Biomolecular Interaction Network Database with accession codes 200014 and 200015. We thank T. Correia for *Drosophila* transformation; N. Baker, R. Haltiwanger, S. Hou, Y. Lin, K. Matsuno, G. Struhl, the Bloomington stock center, and the Developmental Studies Hybridoma Bank for reagents; and J. Culi, R. Haltiwanger, R. Mann, K. Matsuno, and P. Stanley for discussion and comments on the manuscript. T.O. is a fellow (LT00582) of the Human Frontier Science Program. Research in K.D.I.'s laboratory is supported by the Howard Hughes Medical Institute and by NIH grant R01-GM54594.

Regulation of the Polarity Protein Par6 by TGFβ Receptors Controls Epithelial Cell Plasticity

Barish Ozdamar,[1,2]* Rohit Bose,[1,2]* Miriam Barrios-Rodiles,[2] Hong-Rui Wang,[2] Yue Zhang,[2] Jeffrey L. Wrana[1,2]†

The transition of cells from an epithelial to a mesenchymal phenotype is a critical event during morphogenesis in multicellular organisms and underlies the pathology of many diseases, including the invasive phenotype associated with metastatic carcinomas. Transforming growth factor β (TGFβ) is a key regulator of epithelial-to-mesenchymal transition (EMT). However, the molecular mechanisms that control the dissolution of tight junctions, an early event in EMT, remain elusive. We demonstrate that Par6, a regulator of epithelial cell polarity and tight-junction assembly, interacts with TGFβ receptors and is a substrate of the type II receptor, TβRII. Phosphorylation of Par6 is required for TGFβ-dependent EMT in mammary gland epithelial cells and controls the interaction of Par6 with the E3 ubiquitin ligase Smurf1. Smurf1, in turn, targets the guanosine triphosphatase RhoA for degradation, thereby leading to a loss of tight junctions. These studies define how an extracellular cue signals to the polarity machinery to control epithelial cell morphology.

Epithelial-to-mesenchymal transition (EMT) is a striking example of cellular plasticity that involves the dissolution of epithelial tight junctions, modulation of adherens junctions, reorganization of the actin cytoskeleton, loss of apical-basal polarity, and induction of a mesenchymal gene-expression program (*1, 2*). The loss of epithelial homeostasis and the resultant acquisition of a migratory, mesenchymal phenotype are also essential for tumor invasion. Transforming growth factor β (TGFβ) has emerged as a key regulator of EMT in late-stage carcinomas, where it promotes invasion and metastasis (*3, 4*). TGFβ signals through type I and type II transmembrane receptor serine-threonine kinases (TβRI and TβRII, respectively). In the classically described Smad pathway (*5–8*), constitutively active TβRII phosphorylates and activates TβRI, which in turn phosphorylates receptor-regulated Smads. These then bind to Smad4 and accumulate in the nucleus, where they regulate gene transcription by interacting with DNA binding partners. Although both Smad-dependent and Smad-independent mechanisms have pivotal roles in regulating TGFβ-dependent acquisition of a mesenchymal phenotype, the contribution of nontranscriptional signaling conduits to the regulation of epithelial cell plasticity remains less well characterized.

Par6 interacts with the TGFβ receptor. Using LUMIER (*9*), a luminescence-based assay for protein-protein interactions in mammalian cells, we performed a screen to identify interaction partners for TβRI. Briefly, in a 96-well format, human embryonic kidney (HEK293T) cells were transfected with a panel of FLAG-tagged prey proteins in conjunction with TβRI fused to Renilla luciferase (TβRI-Rluc). Lysates were subjected to immunoprecipitation with an antibody to FLAG (anti-FLAG) and assayed for luciferase activity to detect putative interactions with the receptor. In a preliminary screen, known interacting partners of TβRI, such as FKBP12 and TβRI itself as well as serine-threonine kinase receptor–associated protein (STRAP), were confirmed as interactors (Fig. 1A). This screen also identified partitioning-defective protein 6C (Par6C). Par6 is a component of a metazoan polarity complex that was originally characterized in *Caenorhabditis elegans* as a determinant of asymmetric cell cleavage (*10*). More recently, Par6 has been proposed to function as a scaffold for the assembly of a protein complex that regulates cell polarity and includes the Rho guanosine triphosphatase (GTPase) Cdc42, atypical protein kinase C ζ (PKCζ), and the PDZ domain–containing adapter Par3 (*11, 12*). Moreover, Par6 associates with components of additional polarity complexes, including Stardust-Crumbs-PATJ

[1]Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Canada. [2]Program in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

*These authors contributed equally to this work.
†To whom correspondence may be addressed. E-mail: wrana@mshri.on.ca

(13–15) and DLG-SCRIB-LGL (16, 17). Par6 associates with proteins in these complexes through interaction domains that include a PB1 domain, a pseudo-CRIB (Cdc42/Rac interactive binding) motif, and a PDZ domain. Although Par6 regulates the assembly of tight junctions and the apical-basal polarity of mammalian epithelial cells (18), the mechanism for this action is unknown.

TGFβ induces the loss of tight junctions during EMT (19). Therefore, we sought to investigate whether Par6 functions in this pathway. Immunoprecipitation and immunoblotting revealed association of endogenous Par6 with endogenous TβRI in mink lung epithelial (Mv1Lu) cells and demonstrated that Par6 remained bound to TβRI in both the absence and the presence of TGFβ (Fig. 1B). This interaction was specific for TβRI, because we were unable to detect any interaction between wild-type TβRII and Par6 when TβRII was expressed alone (Fig. 1C). We used affinity labeling with [^{125}I]TGFβ to show association of TGFβ receptor complexes from the cell surface with immunoprecipitated Par6, as well as interactions with receptor complexes containing either wild-type or kinase-deficient (KR) TβRII or TβRI (Fig. 1D) (20). Thus, Par6 can associate with the TGFβ receptor complex by interaction with TβRI, independently of receptor kinase activity.

To map the region of Par6 that interacts with the TGFβ receptor, we generated a series of Par6 truncation mutants. The N-terminal 104 amino acids of Par6 were necessary for binding to TβRI (fig. S1A). This region of Par6 contains a PB1 domain, which mediates interaction with the PB1 domain of its partner protein, PKCζ (11, 12). Thus, we explored the relation between Par6 binding to TβRI and PKCζ. A mutant of Par6, Par6(K19A), which interferes with the PKCζ binding surface of the PB1 domain (21), bound to TGFβ receptors in amounts comparable to those of wild-type Par6 (fig. S1B). However, PKCζ association with the receptor complexes was lost in the presence of Par6(K19A). Thus, the N-terminal region of Par6 is necessary for association with the receptors and for PKCζ recruitment to the receptor complex.

TβRI colocalizes with Par6 at tight junctions. Par6 is localized to tight junctions in epithelial cells (18, 22). To investigate the distribution of cell-surface TβRI, we generated a MYC epitope–tagged version of TβRI (MYC-TβRI) in which the tag is placed in the extracellular domain downstream of the signal sequence. We confirmed the functionality of MYC-TβRI by demonstrating that it induced Smad2 phosphorylation, increased Smad transcriptional activity in a TGFβ-regulated manner, and bound [I^{125}]TGFβ in a TβRII-dependent manner (23). We examined the cell-surface distribution of MYC-TβRI and hemagglutinin (HA)–TβRII transiently expressed together in confluent normal murine mammary gland (NMuMG) monolayers by immunofluorescence confocal microscopy (Fig. 1E) (24). These cells establish apical-basal polarity in culture, form tight junctions, and, like many other epithelial cell types, undergo EMT in
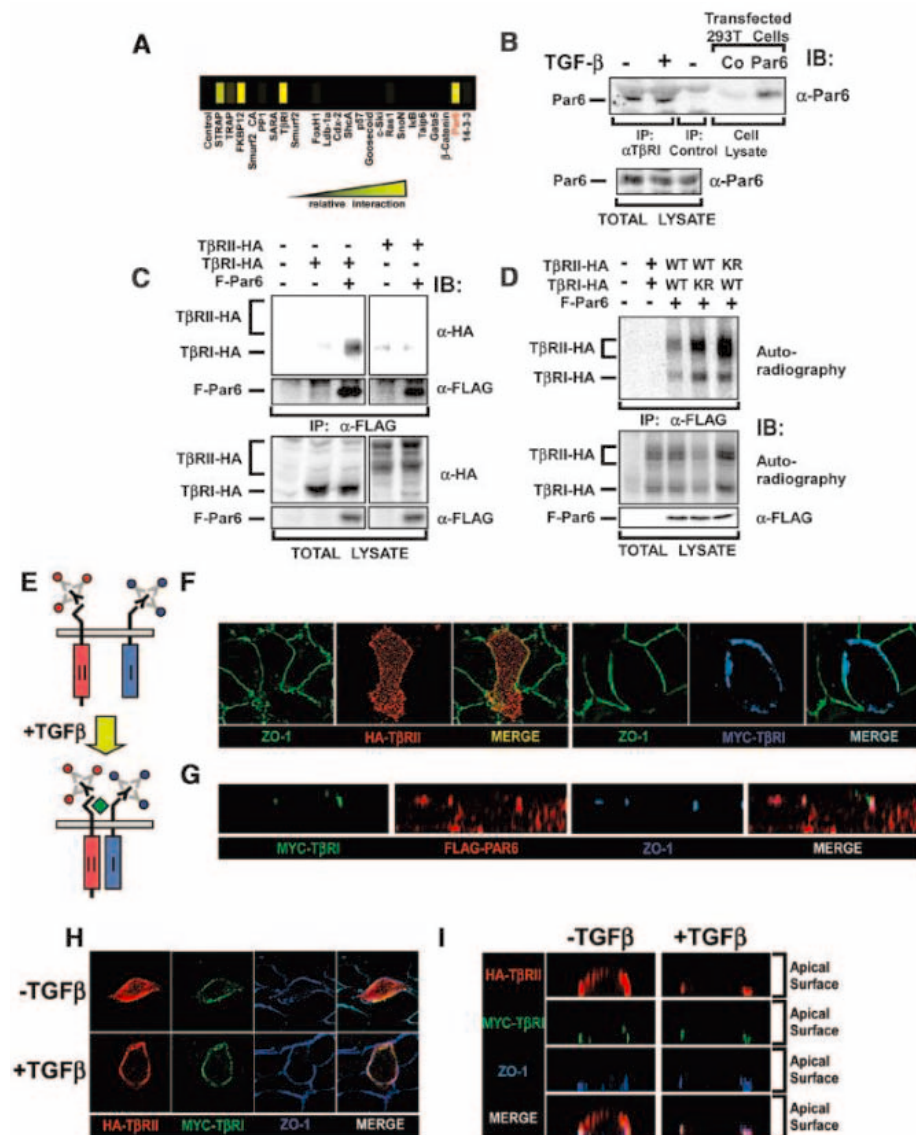
Fig. 1. Binding of Par6 to TβRI. (**A**) LUMIER screen (9) of Renilla luciferase–tagged TβRI against a panel of FLAG-tagged proteins. Yellow represents the relative interaction of TβRI against the indicated protein (9). (**B**) Interaction of endogenous Par6 and TβRI. Lysates from Mv1Lu incubated with (+) or without (–) TGFβ were subjected to immunoprecipitation with antibody to TβRI (α-TβRI) or control antibody (Co) followed by immunoblotting with antibody to Par6 (α-Par6). (**C**) Interaction of Par6 with TβRI but not with TβRII. Lysates from HEK293T cells expressing FLAG-Par6 (F-Par6), TβRI-HA, or TβRII-HA, as indicated, were subjected to anti-FLAG immunoprecipitation (IP) followed by immunoblotting (IB) with anti-FLAG or antibody to HA. Protein expression was confirmed by immunoblotting total lysate. (**D**) Interaction of Par6 with TGFβ receptor complexes at the cell surface. HEK293T cells transfected with Par6 and either wild-type (WT) or KR versions of TβRI and TβRII were affinity labeled, and Par6 bound to receptors was detected by autoradiography, as described (49). Protein amounts were determined as in (B). (**E**) A schematic of the strategy for detecting cell-surface receptors. (**F**) Differential localization of cell-surface TβRI and TβRII in NMuMG cells. HA-TβRII (red) or MYC-TβRI (blue) were detected as described (49), and tight junctions were visualized by ZO-1 staining (green). (**G**) Localization of cell-surface TβRI (green), Par6 (red), and ZO-1 (blue) was determined as described (49). Z-stacks are shown to highlight the tight junctions. (**H**) Localization of cell-surface TβRI (green) and TβRII (red) in tight junctions in NMuMG cells untreated or treated with TGFβ for 30 min, as indicated. (**I**) Z-stacks demonstrate TGFβ-dependent redistribution of TβRII to tight junctions.

response to TGFβ (*19*). Although TβRII was predominantly localized to puncta uniformly distributed over the apical aspect of the cell (Fig. 1F) (*24*, *25*), the surface pool of TβRI was restricted to a discreet band around the apical periphery that colocalized with Zona Occludens-1 (ZO-1) (Fig. 1F), a structural component of tight junctions. Moreover, reconstructed confocal sections along the vertical axis (Z-stack) revealed that cell-surface TβRI localized with ZO-1 on the apical aspect of the cell together with Par6 (Fig. 1G) and was situated apically to endogenous E-cadherin, a marker of adherens junctions (fig. S1C). After stimulation of cells with TGFβ, cell-surface TβRII was redistributed to the tight junctions, where it localized with both TβRI and ZO-1 (Fig. 1, H and I, and movies S1 and S2). TβRI also constitutively interacts with occludin, a structural component of tight junctions, and, upon stimulation of cells with TGFβ, TβRII is recruited to interact with occludin (*11*). Thus, TβRI is localized to tight junctions where Par6 is also found, and TGFβ stimulation induces redistribution of cell-surface TβRII into tight junctions, likely due to ligand-dependent binding of TβRII to TβRI.

Par6 is a substrate of the TGFβ receptor complex. The interaction of Par6 with the TGFβ receptor complex and the localization of cell-surface TβRI to tight junctions suggested that Par6 might serve as a substrate of the receptor complex and be involved in TGFβ-dependent EMT. We analyzed phosphorylation of Par6 bound to endogenous TβRI in [^{32}P]phosphate-labeled NMuMG cells. This revealed TGFβ-dependent phosphorylation of TβRI-bound Par6 in vivo (Fig. 2A). To determine whether Par6 phosphorylation required receptor kinase activity, we analyzed FLAG-Par6 phosphorylation in HEK293T cells that expressed either wild-type or KR receptor complexes. When cells are transfected with both receptors, TβRII and TβRI associate because of intrinsic affinity (*26*) and signal in a kinase activity–dependent manner even in the absence of ligand. In the absence of transfected receptors, phosphorylated Par6 migrated as a single species, whereas in the presence of wild-type receptors we observed a slower migrating form of Par6, which we call Par6* (Fig. 2B). Phosphoamino acid analysis showed that Par6 is phosphorylated on serine residues in the presence of wild-type TGFβ receptors (*23*). We also analyzed the Par6 species separately by two-dimensional (2D) tryptic phosphopeptide mapping. Par6* contained a prominent neutrally charged phosphopeptide that was dependent on the kinase activity of TβRII, as well as several other peptides of weaker intensity (Fig. 2C). Par6* is a distinct phosphoisoform,

as it completely lacked the tryptic phosphopeptide present in the faster migrating form.

To map the site of phosphorylation, we analyzed Par6 mutants. Because truncation

of the C terminus in Par6(1–338) abolished the neutrally charged phosphopeptide (Fig. 2D), we generated serine-to-alanine point mutations at the two serine residues in this peptide and assessed TGFβ receptor–dependent
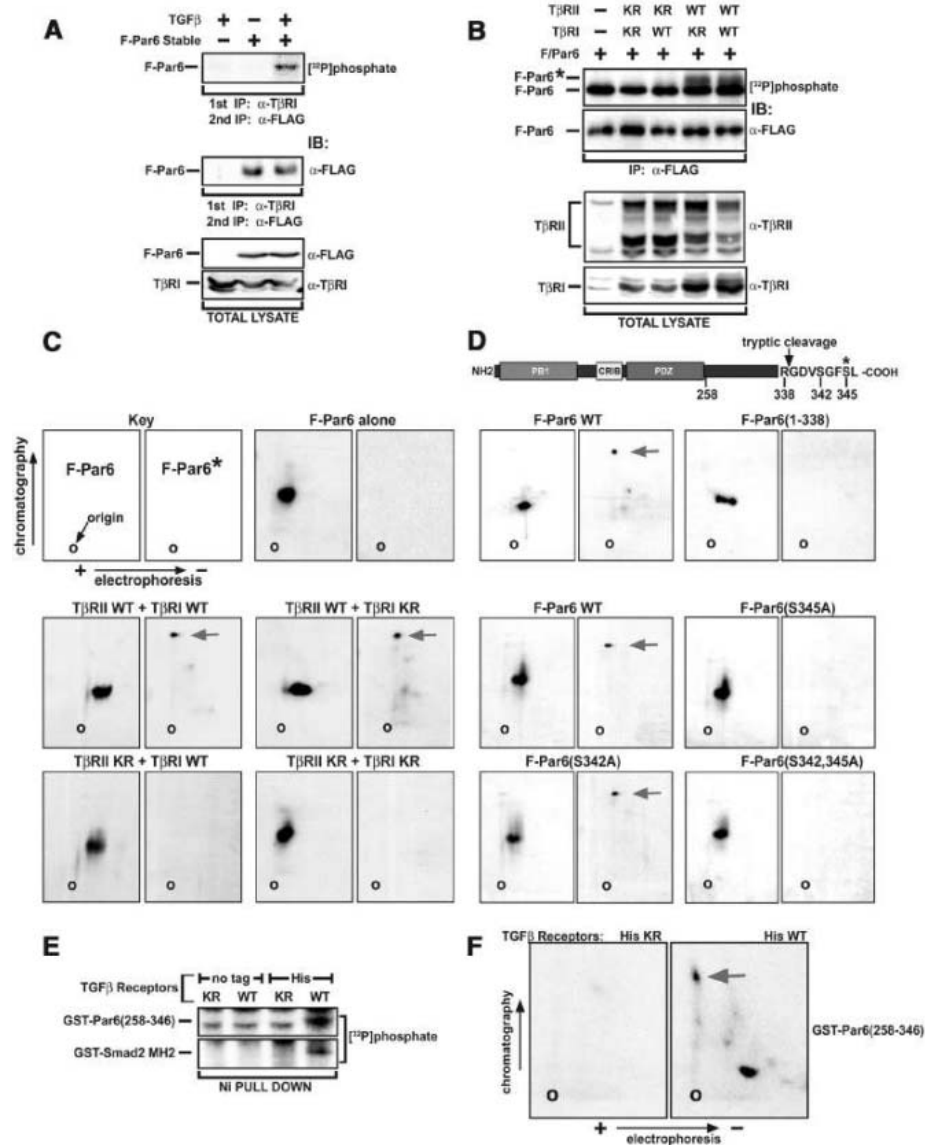
Fig. 2. Par6 as a substrate of the TGFβ receptor complex. (**A**) TGFβ-dependent phosphorylation of Par6. FLAG-tagged Par6-expressing NMuMG cells were [^{32}P]phosphate labeled and treated with (+) or without (−) TGFβ for 1 hour before isolating TGFβ receptor–bound Par6 (*49*). Expression of Par6 and interaction with TβRI was verified by immunoprecipitation and immunoblotting in nonlabeled cells as indicated. (**B**) Dependence of Par6 phosphorylation on TβRII kinase activity. HEK293T cells expressing F-Par6 and the indicated combinations of WT or KR TβRII and TβRI were [^{32}P]phosphate labeled and Par6 isolated by anti-FLAG immunoprecipitation. F-Par6 expression was verified by anti-FLAG immunoblotting of a portion of the immunoprecipitate, and TGFβ receptor expression was verified in nonlabeled cells transfected in parallel. (**C**) 2D tryptic phosphopeptide mapping of Par6. F-Par6 and F-Par6* were isolated from [^{32}P]phosphate-labeled cells expressing TβRI and TβRII as in (B) and were subjected to 2D tryptic phosphopeptide mapping (*49*). The TGFβ receptor–dependent phosphopeptide is indicated by a red arrow. (**D**) Mapping the TGFβ receptor–dependent Par6 phosphorylation site. The indicated mutants of F-Par6 were subjected to tryptic phosphopeptide mapping, as in (C). (**E**) In vitro phosphorylation of Par6 by isolated receptor complexes. HEK293T cells were transiently transfected with the indicated combinations of TβRI and TβRII. Receptor complexes were isolated by nickel purification and used in an in vitro kinase assay with bacterially expressed GST-Par6(258–346) or GST-Smad2 MH2 domain as substrate. (**F**) 2D tryptic phosphopeptide mapping of GST-Par6(258-346). Phosphorylated GST-Par6(258–346) was subjected to phosphopeptide mapping as in (C).

phosphorylation. Whereas both wild-type Par6 and Par6(S342A) were phosphorylated, Par6(S345A) and the Par6(S342, 345A) double mutant were not. We analyzed the phosphorylation of a bacterially expressed C-terminal fragment of Par6 by TGFβ receptors in vitro. TGFβ receptors phosphorylated glutathione S-transferase (GST)–Par6(258–346) in vitro to a similar degree as

they did the MH2 domain of Smad2, a known substrate of the receptor complex (27) (Fig. 2E). Furthermore, 2D tryptic phosphopeptide mapping revealed a phosphopeptide that migrated similarly to one from Par6 phosphorylated by TGFβ receptors in vivo (Fig. 2F). These data suggest that TβRII phosphorylates Par6 at its penultimate residue, Ser[345]. At the C terminus of

Par6 from various metazoans, Ser[345] either is conserved or is a threonine residue in vertebrates, but it is not found in Par6 from *Drosophila* or *C. elegans* (fig. S2A). The latter two species lack tight junctions (28), which suggests that C-terminal phosphorylation of Par6 may be associated with tight-junction homeostasis in vertebrates. Altogether, our results indicate that in unstimulated cells, TβRI and Par6 exist in a complex at tight junctions. TGFβ stimulation then induces assembly of the TβRI-TβRII heteromer, thus bringing the type II receptor to TβRI-bound Par6, where its constitutively active kinase phosphorylates Par6 on Ser[345].

Par6 mutants block TGFβ-induced tight-junction dissolution. Par6 has been proposed to negatively regulate tight junctions (18). Therefore, we sought to determine whether Ser[345] phosphorylation of Par6 was important for TGFβ-dependent dissolution of tight junctions during EMT. We generated multiple NMuMG clonal lines stably expressing comparable amounts of FLAG-tagged wild-type Par6 or Par6(S345A) (fig. S2B) that were correctly localized to tight junctions (fig. S2C). However, although expression of wild-type Par6 did not affect TGFβ-induced EMT, Par6(S345A)-expressing clones displayed stable tight junctions even after prolonged stimulation with TGFβ (Fig. 3A). These clones were also resistant to the TGFβ-induced actin cytoskeletal rearrangement that is associated with the acquisition of a mesenchymal phenotype (Fig. 3A) (19), and they retained cortical actin staining. Similarly, expression of Par6(S345A) blocked the dissolution of adherens junctions in response to TGFβ, as marked by E-Cadherin staining (fig. S2D), as well as the loss of adherens junction–associated structures, as marked by β-Catenin staining (23). Thus, phosphorylation of Par6 by the TGFβ receptor is important for TGFβ-dependent dissolution of tight junctions and rearrangement of the actin cytoskeleton during EMT.

Par6(S345A) does not block Smad activation. To further characterize the EMT block, we examined the transcriptional response of the gene that encodes vimentin, an intermediate filament protein of mesenchymal cells (29). TGFβ stimulated expression of vimentin in Par6(S345A)-expressing cells despite the complete retention of tight junctions (Fig. 3B). Thus, abrogating TGFβ-dependent Par6 phosphorylation interferes with the dissolution of tight junctions but does not impair the induction of vimentin gene expression that is associated with EMT.

Smad-dependent transcriptional regulation is an important mediator of TGFβ-induced EMT, likely participating in the in-
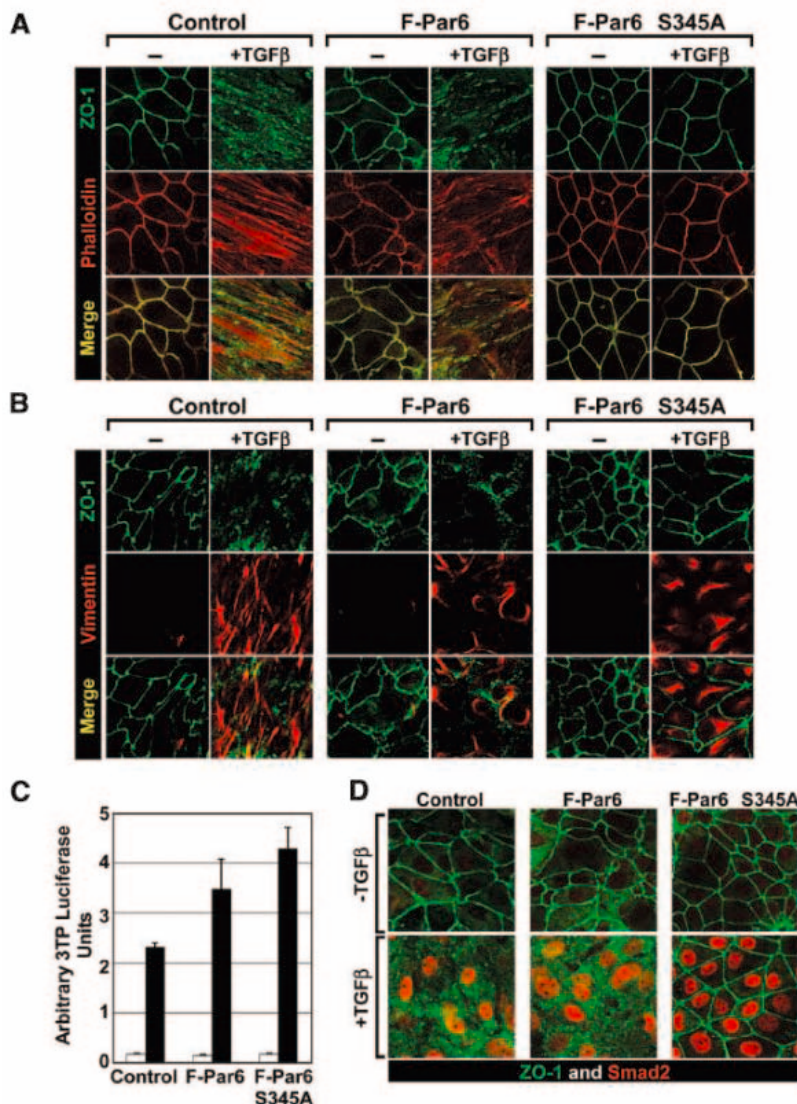
Fig. 3. Phosphorylation of Par6 on Ser[345] is required for TGFβ-dependent EMT in NMuMG cells. (**A**) TGFβ-induced tight-junction dissolution in NMuMG cells. Control NMuMG or clones expressing either Par6 or Par6(S345A), as indicated, were grown to confluent monolayers and then treated with (+) or without (–) TGFβ for 48 hours before staining with Cy3-conjugated phalloidin to detect filamentous actin (red) and immunostaining for ZO-1 (green) (49). Overlay of the two images is shown (merge), with colocalization appearing as yellow. (**B**) Expression of Par6(S345A) does not block induction of vimentin expression. Cells as in (A) were incubated with or without TGFβ for 60 hours before staining with antibody to vimentin (red) and antibody to ZO-1 (green). (**C**) Smad-dependent transcription in NMuMG cells is not impaired by Par6. NMuMG cells were transiently transfected with either control plasmid Par6 or Par6(S345A) in conjunction with 3TP luciferase and were treated with 100 pM TGFβ (closed bars) or left untreated (open bars) for 24 hours before lysis and measurement of luciferase activity. (**D**) Nuclear accumulation of endogenous Smad2 in response to TGFβ stimulation. Control, Par6-, and Par6(S345A)-expressing NMuMG cells were grown to confluence and incubated with or without TGFβ for 48 hours. Localization of endogenous Smad2 was visualized by immunostaining with mouse monoclonal antibody to Smad2 (red) and ZO-1 (green), as described (49).

duction of the mesenchymal gene-expression program (*30*, *31*). Therefore, we sought to determine whether Par6(S345A) hindered Smad signaling. We examined whether transfection of either Par6 or Par6(S345A) in NMuMG cells interfered with expression of the 3TP-luciferase TGFβ reporter gene (*20*), which provides a transcriptional readout of Smad activation. TGFβ induced a strong expression of 3TP-luciferase that was not inhibited by expression of wild-type Par6 or Par6(S345A) (Fig. 3C). Immunofluorescent staining of endogenous Smad2 revealed that, although Par6(S345A) mutant clones retained their tight junctions in response to TGFβ stimulation, nuclear accumulation of Smad2 was comparable to that in control cells at either 4 hours, 24 hours, or 48 hours after stimulation with TGFβ (Fig. 3D) (*23*). Similarly, Smad2 phosphorylation in response to TGFβ was comparable in control, Par6, and Par6(S345A) clones (fig. S2E).

TGFβ regulation of the Par6-Smurf1-RhoA pathway is required for EMT. Smurf1, an E3 ubiquitin ligase, functions as an effector of the polarity complex by mediating localized ubiquitination and degradation of RhoA in cellular protrusions (*32*). RhoA is the prototypical member of the Rho GTPase family, which regulates many cellular processes, including cellular adhesion, motility, and polarity (*33*), and is an important modulator of cell junction formation and stability (*34–36*). Therefore, we hypothesized that in polarized epithelial cells, phosphorylation of Par6 by the TGFβ receptor might regulate tight-junction dissolution by controlling association of Par6 with Smurf1 and the localized degradation of RhoA. We analyzed the interaction of endogenous Smurf1 with either FLAG-Par6 or FLAG-Par6(S345A) in NMuMG cells. In the absence of TGFβ signaling, we detected that some Smurf1 precipitated with wild-type Par6, and the amount increased after treatment of cells with TGFβ (Fig. 4A). In contrast, when we examined Par6(S345A) mutants, there was little if any detectable association with Smurf1. This suggests that Smurf1 preferentially binds Par6 that is phosphorylated by TβRII on Ser³⁴⁵. To confirm this, we compared the phosphorylation pattern of total Par6 to that of Par6 bound to Smurf1 in the presence of TGFβ signaling. Par6 isolated from the total cellular pool displayed tryptic phosphopeptides corresponding to both the faster and slower migrating forms of Par6 (compare Fig. 4B to Fig. 2C). However, when we examined Par6 specifically bound to Smurf1 in the presence of TGFβ signaling, we observed only the phosphorylation pattern that corresponded to the Par6* phosphoisomer, which contains the Ser³⁴⁵ phos-

phopeptide that is a target of TβRII. Hence, TβRII-dependent phosphorylation of Par6 at Ser³⁴⁵ appears to stimulate association with Smurf1.

To investigate the role of Smurf1 in the Par6 pathway, we examined the distribution of endogenous Smurf1 in NMuMG cells. In the absence of signaling, Smurf1 was not
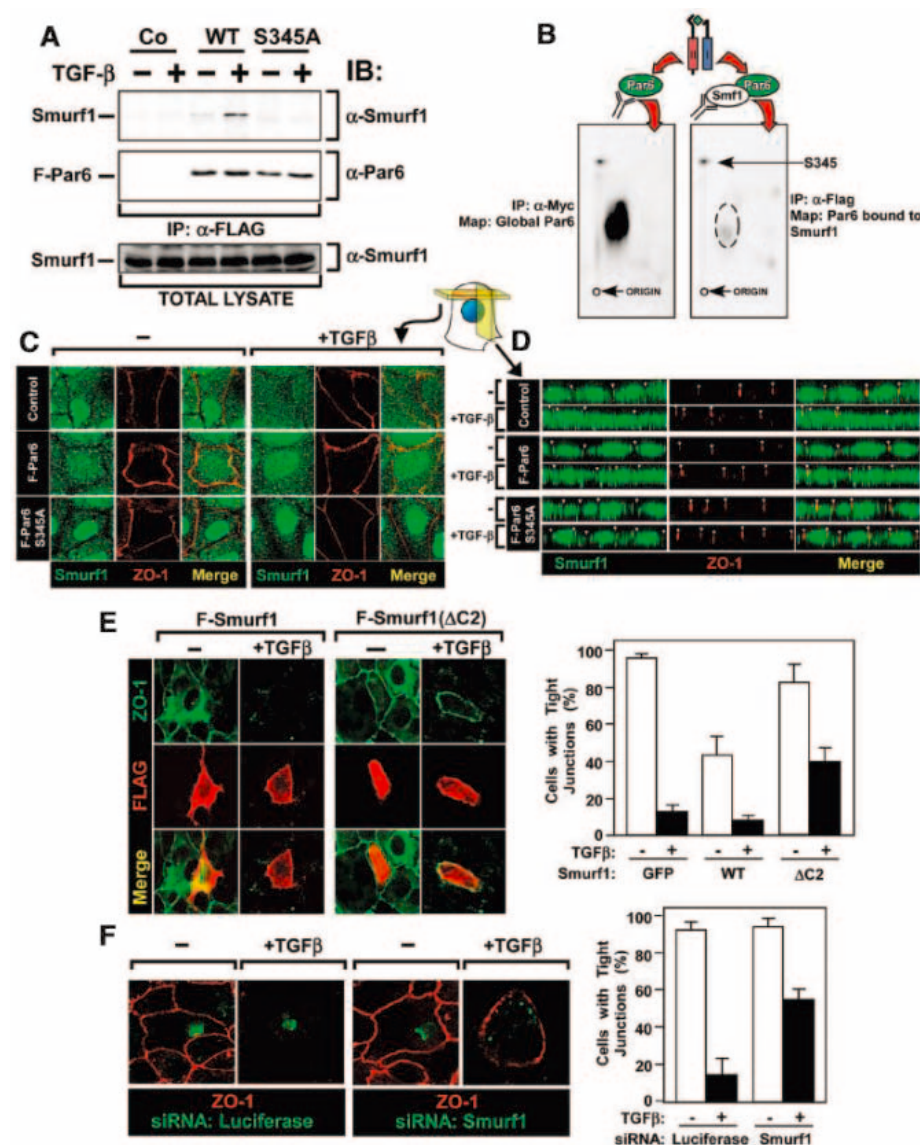
Fig. 4. Interaction of Smurf1 with Par6 phosphorylated on Ser³⁴⁵ and regulation of EMT. (**A**) Assembly of Par6-Smurf1 complexes. Cells expressing FLAG-tagged WT Par6 or Par6(S345A) were treated with (+) or without (−) TGFβ for 1 hour, and lysates were subjected to anti-FLAG immunoprecipitation followed by immunoblotting for Smurf1, as indicated. Protein amounts were confirmed by immunoblotting lysates. (**B**) Par6 bound to Smurf1 is phosphorylated at position Ser³⁴⁵. Total (left map) or FLAG-Smurf1–bound (right map) Par6 was analyzed by phosphopeptide mapping, as in Fig. 2. The position of phosphorylated Ser³⁴⁵ is indicated. (**C**) TGFβ-dependent redistribution of endogenous Smurf1. The indicated NMuMG cells were treated without (−) or with (+) TGFβ for 6 hours before imaging endogenous Smurf1 (green) or ZO-1 (red), as described (*49*). (**D**) Reconstructed Z-stacks derived from confocal optical slices from cells treated as in (C). Colocalization appears yellow, and tight junctions are marked by white arrowheads. (**E**) Smurf1(ΔC2) blocks TGFβ-induced EMT. Confluent NMuMG cells were transiently transfected as indicated and incubated in the absence (−) or presence (+) of TGFβ for 48 hours. Smurf1 was detected by FLAG immunostaining (red) before counterstaining for ZO-1 (green). The percentage of Smurf1-expressing cells with tight junctions in the absence (white) or presence (black) of TGFβ is plotted in the right panel (mean ± SD of 30 cells counted from three experiments). (**F**) Smurf1 siRNA blocks TGFβ-induced EMT. Confluent NMuMG cells were transiently transfected with either FITC-labeled luciferase control siRNA or FITC-labeled Smurf1 siRNA (green) and incubated in the absence (−) or presence (+) of TGFβ for 24 hours before counterstaining for ZO-1 (red). The percentage of Smurf1-siRNA containing cells that retained apical tight junctions was plotted (right panel) (mean ± SD of 30 cells counted from three experiments).

present in ZO-1 labeled regions (Fig. 4, C and D). However, after a 6-hour treatment with TGFβ, a point during EMT when tight junctions have not yet been lost, endogenous Smurf1 redistributed to cell-cell junctions. This TGFβ-dependent redistribution in control cells did not occur in NMuMG cells stably expressing Par6(S345A) (Fig. 4, C and D). To examine the role of Smurf1, we overexpressed either wild-type Smurf1 or Smurf1(ΔC2), which is a mislocalizing mutant that interferes with Smurf1 function

(37). Unlike green fluorescent protein (GFP)–expressing control cells, more than 50% of Smurf1-expressing cells displayed impaired tight-junction stability or were located on the basolateral surface of the epithelial monolayer and exhibited a mesenchymal morphology even in the absence of TGFβ stimulation (Fig. 4E). This morphology is consistent with Smurf1 acting as a factor that promotes mesenchymal transition. In contrast, Smurf1(ΔC2)-expressing cells failed to undergo constitutive EMT and were

also resistant to TGFβ-induced EMT (Fig. 4E). Next, we examined whether small interfering RNA (siRNA) against Smurf1 (35) affected TGFβ-dependent EMT. In cells transfected with control fluorescein isothiocyanate (FITC)–conjugated luciferase siRNA, tight-junction dissolution was unaffected, whereas in cells transfected with FITC-conjugated Smurf1 siRNA, tight junctions were retained in more than 50% of the cells after TGFβ treatment (Fig. 4F). When Smurf1 siRNA was transfected along with a siRNA-resistant version of FLAG-tagged Smurf1 (fig. S3A), cells adopted a mesenchymal phenotype consistent with a rescue of Smurf1 function (fig. S3B).

Localized degradation of RhoA is important for Smurf1-dependent regulation of protrusive activity (35). In control cells, TGFβ caused a reduction of 30% in the amount of RhoA, and treatment of cells with the proteasome inhibitor MG132 reversed the decrease (Fig. 5A and fig. S4A). This TGFβ-dependent decrease in RhoA was not observed in Par6(S345A)-expressing cells (Fig. 5B and fig. S4B), which suggests that TGFβ regulates RhoA levels through Par6. Using mutational analysis, we identified lysines 6 and 7 as the RhoA acceptors for ubiquitin transferred by Smurf1 (fig. S5). We confirmed in a RhoA activation assay that the corresponding RhoA(K6,7R) mutant bound GST-Rhotekin, a RhoA effector that associates specifically with active RhoA, to a similar degree as did wild-type RhoA, thereby demonstrating that the mutation does not affect GTP binding (23). Therefore, we examined whether Smurf1-resistant RhoA affected dissolution of tight junctions in response to TGFβ. Although wild-type RhoA had no effect on EMT, RhoA(K6,7R) inhibited TGFβ-dependent EMT (Fig. 5C). Furthermore, the proteasome inhibitor MG132 also impaired TGFβ-dependent dissolution of tight junctions (23), whereas transfection of FITC-conjugated RhoA siRNA resulted in the disruption of tight junctions in more than 60% of transfected cells (23). Thus, Smurf1-dependent RhoA turnover regulates EMT in response to TGFβ.

Altogether, our results demonstrate a direct link between TGFβ receptors, the polarity complex, and the regulation of tight-junction dissolution during EMT. In this model (Fig. 5D), cell-surface TβRI is localized to tight junctions in polarized epithelial cell sheets (9). TGFβ, which induces association of TβRI with TβRII, leads to the accumulation of receptor complexes in tight junctions, thereby bringing the TβRII receptor kinase to the Par6-TβRI complex. This leads to phosphorylation of Par6, which in turn stimulates binding of Par6 to Smurf1. The Par6-Smurf1 complex then mediates the localized ubiquitination of RhoA to enable
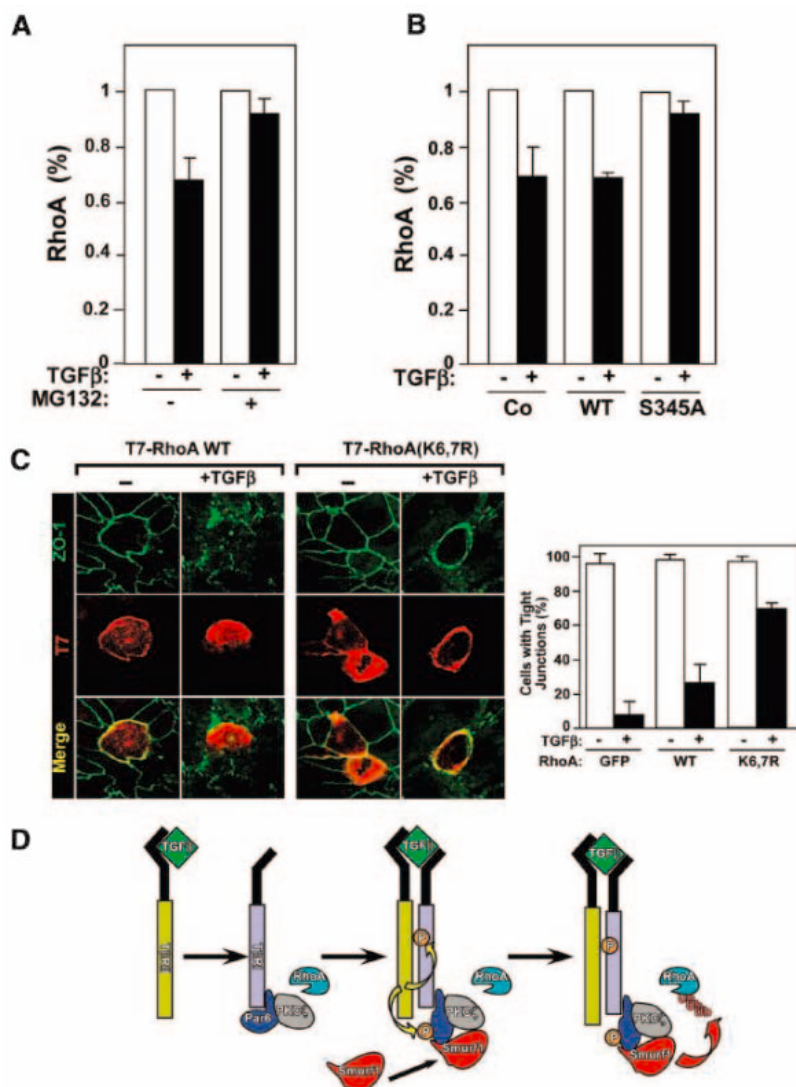
Fig. 5. Requirement of RhoA turnover for TGFβ-dependent EMT. (**A**) TGFβ-dependent decrease in RhoA steady-state levels is blocked by proteasome inhibition. Cells were untreated (white bars) or TGFβ treated (black bars) for 6 hours, either in the absence (–) or presence (+) of the proteasome inhibitor MG132, as indicated. Endogenous RhoA was analyzed by immunoblotting and quantitated on a Fluor-S MultiImager (Bio-Rad, Hercules, CA) and is plotted relative to the amount of RhoA in untreated cells (mean ± SEM of four experiments). (**B**) Par6 (S345A) blocks TGFβ-dependent reduction in RhoA steady-state levels. RhoA steady-state levels in the indicated cell lines untreated (–) or treated (+) with TGFβ were analyzed as in (A). (**C**) Expression of RhoA(K6,7R) blocks TGFβ-dependent EMT. Cells transiently expressing T7 epitope–tagged WT RhoA (T7-RhoA) or mutant RhoA [T7-RhoA(K6,7R)] were analyzed for TGFβ-dependent EMT, as described in Fig. 4D. The percentage of RhoA-expressing cells with tight junctions, in the absence (white bars) or presence (black bars) of TGFβ, was quantitated relative to GFP-expressing control cells (right). (**D**) A model of the TGFβ-Par6 pathway.

the TGFβ-dependent dissolution of tight junctions during EMT.

The transition of epithelia from a highly organized sheet of cells exhibiting apical-basal polarity to a mesenchymal phenotype allows for independent motile behavior that is unrestricted by the 2D layer of the epithelium. In multicellular organisms, this may permit the evolution of specialized form and function during development (2), but in cancer, it is a critical pathological event during progression to a metastatic phenotype (1). EMT is a complex process that involves rearrangements of cell-cell contacts and cell-matrix contacts, as well as changes in gene expression. The actin cytoskeleton is extensively remodeled from its cortical distribution in epithelia to form contractile stress fibers typical of motile fibroblasts. However, it has been unclear how these events are orchestrated by extracellular cues such as TGFβ. In epithelial cells, polarity protein complexes have been proposed to serve as important modulators of tight-junction homeostasis and apical-basal polarity (13). Par6 is a key component of this machinery and functions as a scaffold to regulate the recruitment of effectors. Here, we link TGFβ signaling directly to the Par6-Smurf1 polarity complex. We previously described a PKCζ-dependent role for Smurf1 and localized RhoA turnover in the formation of cellular protrusions (32). It is unclear whether the mechanisms that regulate TGFβ-dependent assembly of the Smurf1-Par6 complex in cell junctions differ from those that control protrusion dynamics. However, integrin signaling can regulate the Par6 complex to control directed cell motility (38), and receptor tyrosine kinases are required in Par6 modulation of axonal polarity (39). Although the molecular mechanisms underlying these pathways are unclear, they suggest that Par6 may be a focal point for the control of cell polarity by multiple extracellular cues not only during development and homeostasis but also in the progression of cancer.

Reorganization of the cytoskeleton, and in particular the formation of cortical actin filaments, is critical for the establishment of tight and adherens junctions and the regulation of epithelial cell apical-basal polarity. The concerted activity of multiple Rho GTPases, which are key regulators of the actin cytoskeleton, coordinates these events (35). Notably, RhoA functions to maintain apical-basal polarity and cell junctions in colonic epithelia and keratinocytes (34, 36), and several reports have suggested that RhoA can stabilize tight junctions and increase transepithelial resistance (40–42). These observations are consistent with our findings, as we propose that the localized degradation of RhoA is important in TGFβ-mediated EMT. However, other evidence suggests that RhoA activity promotes TGFβ-induced EMT (43, 44) and stress-fiber formation (33). These divergent outputs may be explained in part by the activity of downstream effectors of RhoA. RhoA-induced Rho-kinase (ROCK) activity can lead to the dissolution of cell-cell contacts and the formation of stress fibers, whereas mDia, another downstream effector of RhoA, promotes junction stabilization (36). Consequently, how RhoA regulates junction dynamics may be distinct from the mechanism underlying formation of stress fibers. Inhibition of ROCK blocks TGFβ-dependent formation of stress fibers but not dissolution of tight junctions (45); therefore, the requirement for RhoA activity in TGFβ-induced EMT may reflect its function in the latter pathway. Thus, RhoA may fulfill multiple spatiotemporal roles during TGFβ-induced EMT.

In late-stage cancers, autocrine production of TGFβ induces EMT and synergizes with activated Ras to promote metastasis (46). A number of other signaling conduits, including Smads, the mitogen-activated protein kinase p38, and phosphoinositide-3 kinase, have also been implicated in this process (3, 31, 47, 48). The specific events that are regulated by the latter two pathways are unclear; however, Smads are critical for the transcriptional responses accompanying EMT. Thus, our demonstration that both Smad activation and the induction of mesenchymal gene expression can be dissociated from a pathway controlling loss of tight junctions reveals a bifurcated signaling network at the level of the TGFβ receptor. This dynamic interplay of multiple pathways likely allows the TGFβ receptor system to orchestrate EMT by coordinating dissolution of tight junctions with gene-expression programs. Moreover, our high-throughput systematic analysis of protein-protein interactions in mammalian cells revealed that the TGFβ-Par6 signaling pathway is embedded in a larger network within which TGFβ components extensively connect to the p21-activated kinase signaling network (9). It will be interesting to determine the biological contexts in which this larger network functions to interpret TGFβ-family signaling.

References and Notes
1. S. Grunert, M. Jechlinger, H. Beug, Nature Rev. Mol. Cell Biol. 4, 657 (2003).
2. D. Shook, R. Keller, Mech. Dev. 120, 1351 (2003).
3. P. M. Siegel, J. Massague, Nature Rev. Cancer 3, 807 (2003).
4. M. Oft, K. H. Heider, H. Beug, Curr. Biol. 8, 1243 (1998).
5. L. Attisano, J. L. Wrana, Science 296, 1646 (2002).
6. Y. Shi, J. Massague, Cell 113, 685 (2003).
7. R. Derynck, Y. E. Zhang, Nature 425, 577 (2003).
8. A. Moustakas, S. Souchelnytskyi, C. H. Heldin, J. Cell Sci. 114, 4359 (2001).
9. M. Barrios-Rodiles et al., Science 307, 1621 (2005).
10. J. L. Watts et al., Development 122, 3133 (1996).
11. D. Lin et al., Nature Cell Biol. 2, 540 (2000).
12. G. Joberty, C. Petersen, L. Gao, I. G. Macara, Nature Cell Biol. 2, 531 (2000).
13. T. W. Hurd, L. Gao, M. H. Roh, I. G. Macara, B. Margolis, Nature Cell Biol. 5, 137 (2003).
14. S. C. Nam, K. W. Choi, Development 130, 4363 (2003).
15. C. Lemmers et al., Mol. Biol. Cell 15, 1324 (2004).
16. P. J. Plant et al., Nature Cell Biol. 5, 301 (2003).
17. J. Betschinger, K. Mechtler, J. A. Knoblich, Nature 422, 326 (2003).
18. L. Gao, G. Joberty, I. G. Macara, Curr. Biol. 12, 221 (2002).
19. P. J. Miettinen, R. Ebner, A. R. Lopez, R. Derynck, J. Cell Biol. 127, 2021 (1994).
20. J. L. Wrana et al., Cell 71, 1003 (1992).
21. Y. Noda et al., J. Biol. Chem. 278, 43516 (2003).
22. A. Johansson, M. Driessens, P. Aspenstrom, J. Cell Sci. 113, 3267 (2000).
23. B. Ozdamar, R. Bose, unpublished data.
24. G. M. Di Guglielmo, C. Le Roy, A. F. Goodfellow, J. L. Wrana, Nature Cell Biol. 5, 410 (2003).
25. M. Ehrlich, A. Shmuely, Y. I. Henis, J. Cell Sci. 114, 1777 (2001).
26. X. H. Feng, R. Derynck, J. Biol. Chem. 271, 13123 (1996).
27. M. Macias-Silva et al., Cell 87, 1215 (1996).
28. E. Willott et al., Proc. Natl. Acad. Sci. U.S.A. 90, 7834 (1993).
29. R. Pagan, I. Martin, A. Alonso, M. Llobera, S. Vilaro, Exp. Cell Res. 222, 333 (1996).
30. A. Moustakas, K. Pardali, A. Gaal, C. H. Heldin, Immunol. Lett. 82, 85 (2002).
31. A. V. Bakin, C. Rinehart, A. K. Tomlinson, C. L. Arteaga, J. Cell Sci. 115, 3193 (2002).
32. H. R. Wang et al., Science 302, 1775 (2003).
33. K. Burridge, K. Wennerberg, Cell 116, 167 (2004).
34. A. Vaezi, C. Bauer, V. Vasioukhin, E. Fuchs, Dev. Cell 3, 367 (2002).
35. M. Perez-Moreno, C. Jamora, E. Fuchs, Cell 112, 535 (2003).
36. E. Sahai, C. J. Marshall, Nature Cell Biol. 4, 408 (2002).
37. C. Suzuki et al., J. Biol. Chem. 277, 39919 (2002).
38. S. Etienne-Manneville, A. Hall, Cell 106, 489 (2001).
39. S. H. Shi, L. Y. Jan, Y. N. Jan, Cell 112, 63 (2003).
40. H. Fujita et al., Biochem. J. 346, 617 (2000).
41. S. A. Rajasekaran et al., Mol. Biol. Cell 12, 3717 (2001).
42. M. L. Chen, C. Pothoulakis, J. T. LaMont, J. Biol. Chem. 277, 4247 (2002).
43. N. A. Bhowmick et al., Mol. Biol. Cell 12, 27 (2001).
44. Y. C. Tian, D. Fraser, L. Attisano, A. O. Phillips, Am. J. Physiol. Renal Physiol. 285, F130 (2003).
45. V. Kaartinen, L. Haataja, A. Nagy, N. Heisterkamp, J. Groffen, Int. J. Mol. Med. 9, 563 (2002).
46. E. Janda et al., J. Cell Biol. 156, 299 (2002).
47. A. V. Bakin, A. K. Tomlinson, N. A. Bhowmick, H. L. Moses, C. L. Arteaga, J. Biol. Chem. 275, 36803 (2000).
48. L. Yu, M. C. Hebert, Y. E. Zhang, EMBO J. 21, 3749 (2002).
49. Materials and methods are available as supporting material on Science Online.

REPORTS

The Magnetic Field of the Large Magellanic Cloud Revealed Through Faraday Rotation

B. M. Gaensler,[1,2]* M. Haverkorn,[1] L. Staveley-Smith,[3]
J. M. Dickey,[4] N. M. McClure-Griffiths,[3] J. R. Dickel,[5] M. Wolleben[6]

We have measured the Faraday rotation toward a large sample of polarized radio sources behind the Large Magellanic Cloud (LMC) to determine the structure of this galaxy's magnetic field. The magnetic field of the LMC consists of a coherent axisymmetric spiral of field strength ∼1 microgauss. Strong fluctuations in the magnetic field are also seen on small (<0.5 parsec) and large (∼100 parsecs) scales. The large bursts of recent star formation and supernova activity in the LMC argue against standard dynamo theory, adding to the growing evidence for rapid field amplification in galaxies.

The Milky Way and many other spiral galaxies show well-organized, large-scale magnetic fields (*1–3*), whose existence points to a powerful and ubiquitous process that organizes random motions into coherent magnetized structures. The underlying mechanism is believed to be a dynamo, in which magnetic fields are slowly ordered and amplified due to the interplay between turbulence and differential rotation (*4, 5*). Magnetism in galaxies is usually mapped by observing the orientation of polarized optical and radio emission from the galaxy itself (*3*), but these data have limited spatial resolution (*6*) and can be difficult to interpret (*7*). An alternative is the direct determination of the geometry and strength of magnetic fields, which comes from the Faraday rotation of background radio sources (*8–10*), an effect in which birefringence in an intervening magneto-ionized source rotates the plane of linearly polarized radiation. Measurements of background rotation measures (RMs) are free from the difficulties associated with studying polarized emission produced by the source itself, which suffer from a complicated combination of internal and external Faraday rotation, depolarization, and optical extinction.

We studied the magnetic field of the Large Magellanic Cloud (LMC) using 1.4-GHz polarization data recorded as part of a hydrogen line survey (*11*) carried out with the Australia Telescope Compact Array (*12*). Over a field of 130 square degrees, we calculated RMs for 291 polarized background sources. Of this sample, about 100 sources lie directly behind the LMC. We used 140 measurements of sources lying outside the LMC to subtract a mean RM, presumably resulting from foreground Faraday rotation in the Milky Way. The resulting distribution of residual Faraday rotation (Fig. 1) shows a strong excess in RM across the extent of the LMC. The implied magnetic field demonstrates spatial coherence: The eastern half of the galaxy shows predominantly positive RMs, and in the west, the RMs are mainly negative.

We converted position on the sky to a location within the LMC disk for each RM measurement, assuming that the galaxy is inclined to the plane of the sky at an angle $i = 35°$, with its line of nodes at a position angle $\Theta = 123°$ (measured north through east) (*13*). The resulting dependence of RM against position angle within the disk of the LMC (Fig. 2) shows a systematic variation, with the maximum mean RM occurring near the line of nodes. These data can be well fit by a cosinusoid of amplitude $RM_0 = 53 \pm 3$ rad m^{-2}, demonstrating that the coherent component of the LMC's magnetic field has an axisymmetric spiral geometry (*14*), as is seen in other galaxies (*2, 3, 14*). The phase of this cosinusoid corresponds to the pitch angle p of the spiral field (*14*), but in this case we can only infer a weak constraint, $|p| \lesssim 20°$, especially given that the uncertainty on Θ is $\gtrsim 10°$ (*13*).

In addition to the coherent field, a structure function analysis indicates random fluctuations in RM, with a standard deviation

$\Sigma_{RM} = 81$ rad m^{-2} and occurring on a characteristic angular scale of ∼0.1°, or a length scale $L \sim 90$ pc at the distance to the LMC of 50 kpc. This may represent the evolved supernova remnants (SNRs) and wind bubbles whose interlocking shells dominate the morphology of ionized gas in the LMC on this scale (*15*).

To estimate the relative strength of the ordered and random field components, we assumed that ionized gas in the LMC consists of a disk of projected thickness D, in which cells of linear size L contain clumps of ionized gas of filling factor f and density n_e (*16*). In each cell, the magnetic field is composed of a uniform component of strength B_0 plus a randomly oriented component of strength B_R. If B_R is uncorrelated with fluctuations in n_e, it can be shown that

$$\frac{\Sigma_{RM}}{|RM_0|} \approx \left[\frac{L}{2fD} \left(1 + \frac{2}{3} \frac{B_R^2}{B_0^2 \sin^2 i} \right) \right]^{1/2} \quad (1)$$

The occupation length fD of ionized gas is $DM_0^2/EM_0 \approx 530$ pc, where $DM_0 \equiv \int n_e \, dl \approx 100$ cm^{-3} pc and $EM_0 \equiv \int n_e^2 \, dl \approx 19$ pc cm^{-6} are the average dispersion measure (DM) and median extinction-corrected emission measure (EM) integrated through the LMC, respectively (*12*). Equation 1 then implies $B_R/B_0 = 3.6$. The random field dominates the ordered field, as seen in many other galaxies (*2, 17*).

The actual values of B_0 and B_R can also be estimated. If B_R and fluctuations in n_e are uncorrelated, then the strength of the ordered component of the LMC's magnetic field is

$$B_0 = \frac{|RM_0|}{K \, DM_0 \sin i} \approx 1.1 \; \mu G \quad (2)$$

where $K = 0.81$ rad m^{-2} pc^{-1} cm^3 μG^{-1}. The strength of the random field is then $B_R = 3.6B_0 \approx 4.1 \; \mu$G, and the total magnetic field strength on large scales is $B_T = (B_0^2 + B_R^2)^{1/2} \approx 4.3 \; \mu$G. We note that in selected regions where B and n_e are correlated, as might result from compression in SNR shocks, the above approach overestimates B_0 and underestimates B_R, each by factors of ∼2 (*17*).

The polarized background sources are not randomly distributed across the LMC's extent but tend to avoid areas of bright Hα emission (Fig. 1). Specifically, background sources are increasingly depolarized as they propagate through regions of higher EM (Fig. 3). This suggests that we are observing beam depolarization, in which small-scale fluctuations in foreground RM produce interference between polarized rays along adjacent sightlines (*7*). Beam depolarization is only important

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Mail Stop 6, Cambridge, MA 02138, USA. [2]School of Physics, University of Sydney, New South Wales 2006, Australia. [3]Australia Telescope National Facility, Commonwealth Scientific and Industrial Research Organization, Post Office Box 76, Epping, New South Wales 1710, Australia. [4]Physics Department, University of Tasmania, GPO Box 252-21, Hobart, Tasmania 7001, Australia. [5]Astronomy Department, University of Illinois, 1002 West Green Street, Urbana, IL 61801, USA. [6]Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany.

*To whom correspondence should be addressed. E-mail: bgaensler@cfa.harvard.edu

when the angular scale of RM fluctuations is smaller than the resolution of the data and the scale of intrinsic polarized structures. Although the resolution here, ~40 arc sec, is comparatively large, extragalactic sources are typically of much smaller angular extent: For 1.4-GHz flux densities in the range 10 to 300 millijanskys (mJy) as observed here, the median angular size is only ~6 arc sec (18), or ~1.5 pc when projected against the LMC. The depolarization (Fig. 3) implies strong RM fluctuations on scales $l \ll 1.5$ pc. In such a situation, beam depolarization reduces the intrinsic linearly polarized intensity, P_0, to a level

$$P = P_0 \exp(-2\sigma_{RM}^2 \lambda^4) \qquad (3)$$

(7, 19), where λ is the observing wavelength and σ_{RM} is the standard deviation in RMs across the source. To account for the dependence of P/P_0 on EM (Fig. 3), we need to relate σ_{RM} to the EM along a given sightline. If the magnetic field is uncorrelated with the ionized gas density, we expect that $\sigma_{RM} = k\mathrm{EM}^{1/2}$, where k is a constant. With this assumption, the fluctuating magnetic field on a scale of l pc needed to produce the observed depolarization has a strength $B_r \approx k(lK^2/3)^{-1/2}$ µG. In Fig. 3, Eq. 3 has been fit to the data for $P_0 \approx 0.104$ and $k \approx 1.8$ rad m^{-2} pc$^{-1/2}$ cm^3. Assuming $l <$ 0.5 pc, we find that $B_r > 5$ µG. We thus infer that there are large RM and magnetic field fluctuations on subparsec scales in the ionized gas of the LMC. This phenomenon is also seen in our own Galaxy and may trace the turbulent winds and H II regions of individual stars (20, 21).

Most spiral galaxies are long-lived systems that exhibit large amounts of rotational shear and that experience relatively constant star formation rates over long periods of time. Coherent magnetic fields in these galaxies are believed to be produced by a dynamo mechanism, in which small-scale turbulent magnetic fields are amplified and ordered by cyclonic motions and differential rotation (2, 4, 5). However, in galaxies dominated by sudden bursts of star formation and supernova activity, the dramatic injection of energy should disrupt the slow monotonic increase of the large-scale field produced by a standard turbulent dynamo (22, 23). The LMC has experienced several intense bursts of star formation over the past ~4 billion years triggered by repeated close encounters with the Milky Way and with the Small Magellanic Cloud (24, 25) and yet still maintains a coherent spiral magnetic field. Combined with previous results demonstrating the presence of ordered magnetic fields in young galaxies for which the dynamo has had little time to operate (26), and in irregular galaxies, which lack large amounts of rotation (27), there is now evidence that standard dynamo processes are ineffective in the LMC and these other galaxies. There are several

viable alternatives to explain the coherent magnetic fields that we observe. Potentially most pertinent for the LMC is the cosmic ray–driven dynamo, in which recent supernova activity generates a large population of relativistic particles. The buoyancy of these
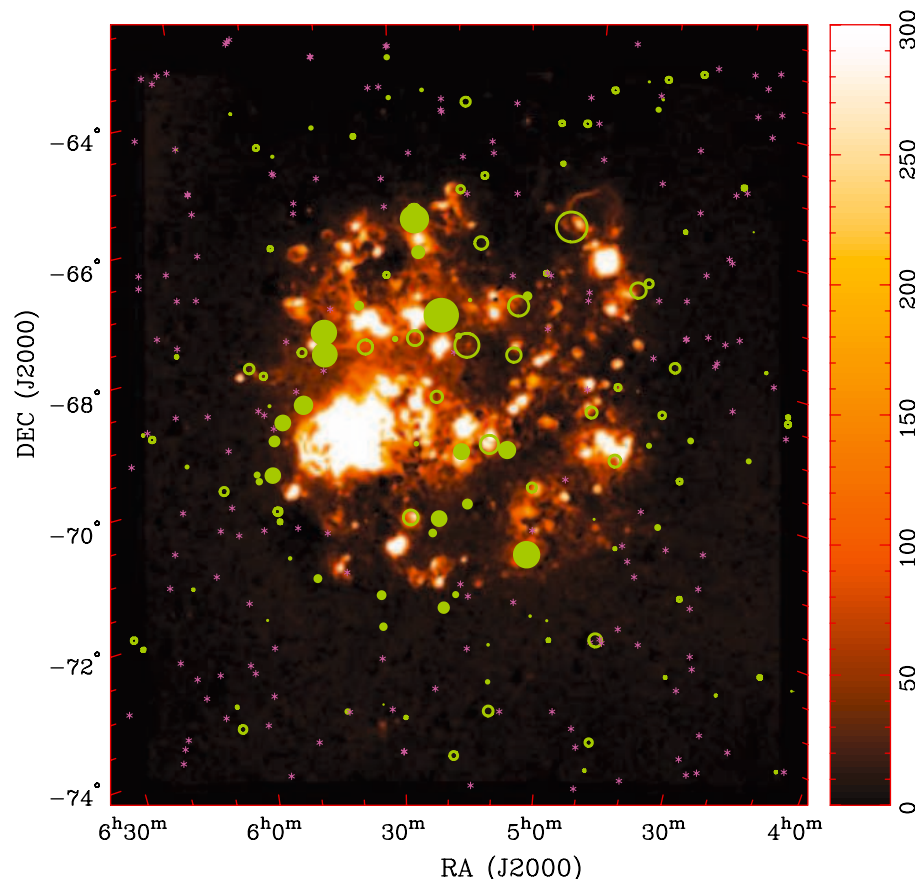
Fig. 1. Faraday rotation measures through the LMC. The image shows the distribution of emission measure toward the LMC in units of pc cm^{-6}, derived from the Southern H-Alpha Sky Survey Atlas (31). The symbols show the position, sign, and magnitude of the baseline-subtracted RM measurements (12). Solid and open circles (both marked in green) correspond to positive and negative RMs, respectively, and asterisks (marked in purple) indicate RMs that are consistent with zero within their errors. The diameter of each circle is proportional to the magnitude of the RM, the largest positive and negative RMs being +247 ± 13 rad m^{-2} and −215 ± 32 rad m^{-2}, respectively. RA, right ascension; DEC, declination.

Fig. 2. RM against position angle (PA) within the LMC. The six data points are a binned representation of the 93 RMs that lie within a radius of 3.5° of the center of the EM distribution seen in Fig. 1 [right ascension (J2000) 05^{h}16^{m}03^s, declination (J2000) −68°41′45″], plotted against deprojected position angle within the LMC, measured from the line of nodes. The uncertainty on each datum is the weighted standard error in the mean for RMs in that bin. The dashed line shows a cosinusoidal least-squares fit to the unbinned data, with an amplitude of +53 ± 3 rad m^{-2} and an offset from zero of +9 ± 2 rad m^{-2}. The phase of the cosinusoid is only weakly constrained, falling between ±15°. The fit is not a strong function of the center adopted for the LMC.
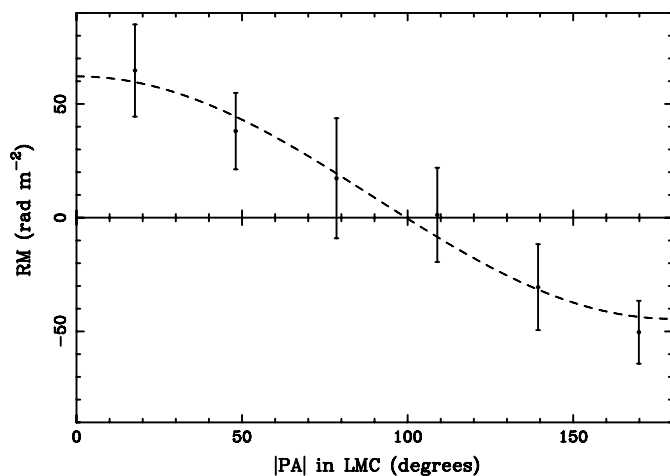
Fig. 3. Polarized fraction of 81 background sources as a function of EM (12 of the 93 sources shown in Fig. 2 have been excluded; six with EM > 100 pc cm^{-6} and six with an observed EM ≤ 0 pc cm^{-6} due to imperfect star subtraction). A Galactic foreground contribution of 3 pc cm^{-6} has been subtracted from each EM measurement. The uncertainty for each binned data point corresponds to the weighted standard error in the mean for each bin. The observed depolarization as a function of EM cannot be a result of source confusion or other observational selection effects because sources with RMs were identified from an image of linear polarization, in which the weak signals from diffuse polarized emission show no correlation with Hα emission. It also cannot be due to excessive Faraday rotation across our observing band (bandwidth depolarization), because for the narrow frequency channels (8 MHz) used here, this effect would manifest itself only for |RM| > 4000 rad m^{-2}, ~20 times the size of any RMs observed. The dashed line shows a least-squares fit of Eq. 3 to the unbinned data, assuming $\sigma_{RM} \propto EM^{1/2}$.
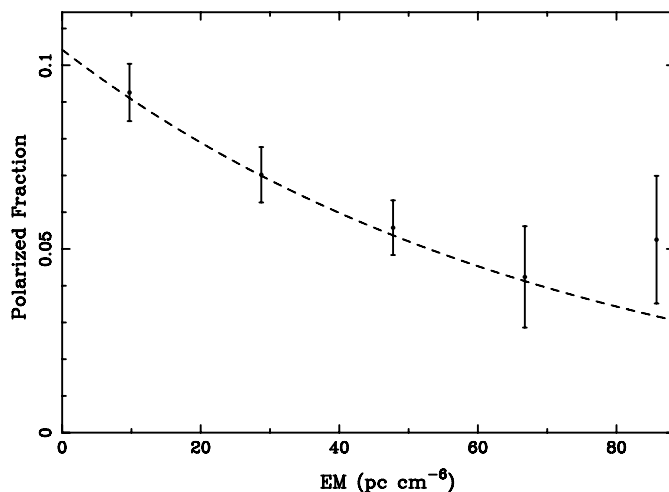
particles inflates magnetic loops out of the disk; adjacent loops reconnect and then are amplified by differential rotation to generate a large-scale spiral field (*28*, *29*). This mechanism not only requires vigorous star formation, as has occurred recently for the LMC, but has a time scale for amplification of only ~0.2 billion years (*29*) and so can quickly generate large-scale magnetic fields before they are dissipated by outflows and tidal interactions. This process can thus potentially account for the coherent fields seen in the LMC and other galaxies (*30*).

that the observed RMs do not probe ionized gas in bright individual H ɪɪ regions, making this a reasonable assumption.

References and Notes

1. R. Beck, A. Brandenburg, D. Moss, A. Shukurov, D. Sokoloff, *Annu. Rev. Astron. Astrophys.* **34**, 155 (1996).
2. R. Beck, *Philos. Trans. R. Soc. London Ser. A* **358**, 777 (2000).
3. J.-L. Han, R. Wielebinski, *Chin. J. Astron. Astrophys.* **2**, 293 (2002).
4. A. A. Ruzmaikin, D. D. Sokolov, A. M. Shukurov, *Magnetic Fields of Galaxies* (Kluwer, Dordrecht, Netherlands, 1988).
5. R. M. Kulsrud, *Annu. Rev. Astron. Astrophys.* **37**, 37 (1999).
6. A. Fletcher, E. M. Berkhuijsen, R. Beck, A. Shukurov, *Astron. Astrophys.* **414**, 53 (2004).
7. D. D. Sokoloff *et al.*, *Mon. Not. R. Astron. Soc.* **299**, 189 (1998).
8. J. C. Brown, A. R. Taylor, B. J. Jackel, *Astrophys. J. Suppl.* **145**, 213 (2003).
9. J. L. Han, R. Beck, E. M. Berkhuijsen, *Astron. Astrophys.* **335**, 1117 (1998).
10. B. M. Gaensler, R. Beck, L. Feretti, *N. Astron. Rev.* **48**, 1003 (2004).
11. S. Kim *et al.*, *Astrophys. J.* **503**, 674 (1998).
12. Materials and methods are available as supporting material on *Science* Online.
13. R. P. van der Marel, *The Local Group as an Astrophysical Laboratory*, M. Livio, Ed. (Cambridge Univ. Press, Cambridge, in press) (arxiv.org/abs/astro-/0404192).
14. M. Krause, E. Hummel, R. Beck, *Astron. Astrophys.* **217**, 4 (1989).
15. J. Meaburn, *Mon. Not. R. Astron. Soc.* **192**, 365 (1980).
16. The depolarization demonstrated in Fig. 3 implies
17. R. Beck, A. Shukurov, D. Sokoloff, R. Wielebinski, *Astron. Astrophys.* **411**, 99 (2003).
18. R. A. Windhorst, E. B. Fomalont, R. B. Partridge, J. D. Lowenthal, *Astrophys. J.* **405**, 498 (1993).
19. B. J. Burn, *Mon. Not. R. Astron. Soc.* **133**, 67 (1966).
20. J. P. Leahy, *Mon. Not. R. Astron. Soc.* **226**, 433 (1987).
21. M. Haverkorn, B. M. Gaensler, N. M. McClure-Griffiths, J. M. Dickey, A. J. Green, *Astrophys. J.* **609**, 776 (2004).
22. P. P. Kronberg, *Rep. Prog. Phys.* **57**, 325 (1994).
23. K. T. Chyży, R. Beck, *Astron. Astrophys.* **417**, 541 (2004).
24. E. W. Olszewski, N. B. Suntzeff, M. Mateo, *Annu. Rev. Astron. Astrophys.* **34**, 511 (1996).
25. K. Bekki, M. Chiba, *Mon. Not. R. Astron. Soc.* **356**, 680 (2005).
26. P. P. Kronberg, J. J. Perry, E. L. H. Zukowski, *Astrophys. J.* **387**, 528 (1992).
27. K. T. Chyży, R. Beck, S. Kohle, U. Klein, M. Urbanik, *Astron. Astrophys.* **355**, 128 (2000).
28. D. Moss, A. Shukurov, D. Sokoloff, *Astron. Astrophys.* **343**, 120 (1999).
29. M. Hanasz, G. Kowal, K. Otmianowska-Mazur, H. Lesch, *Astrophys. J.* **605**, L33 (2004).
30. K. Otmianowska-Mazur, K. T. Chyży, M. Soida, S. von Linden, *Astron. Astrophys.* **359**, 29 (2000).
31. J. E. Gaustad, P. R. McCullough, W. Rosing, D. Van Buren, *Publ. Astron. Soc. Pac.* **113**, 1326 (2001).
32. We thank S. Kim for carrying out the original Australian Telescope Compact Array observations that made this project possible and R. Beck, R. Crutcher, K. Otmianowska-Mazur, D. Elstner, and D. Sokoloff for useful discussions. The Southern H-Alpha Sky Survey Atlas is supported by NSF. The Australia Telescope is funded by the Commonwealth of Australia for operation as a National Facility managed by the Commonwealth Scientific and Industrial Research Organization. Supported by NSF through grant AST-0307358 and by the Denison Fund of the University of Sydney (B.M.G.).

Molecular Mechanisms for the Functionality of Lubricant Additives

Nicholas J. Mosey,[1] Martin H. Müser,[2]* Tom K. Woo[1]

Wear limits the life-span of many mechanical devices with moving parts. To reduce wear, lubricants are frequently enriched with additives, such as zinc phosphates, that form protective films on rubbing surfaces. Using first-principles molecular dynamics simulations of films derived from commercial additives, we unraveled the molecular origin of how antiwear films can form, function, and dissipate energy. These effects originate from pressure-induced changes in the coordination number of atoms acting as cross-linking agents to form chemically connected networks. The proposed mechanism explains a diverse body of experiments and promises to prove useful in the rational design of antiwear additives that operate on a wider range of surface materials, with reduced environmental side effects.

Whenever two surfaces slide past one another, the potential for the deterioration of one or both of these surfaces exists. Although wear is not necessarily an undesirable effect,

such as in manufacturing or polishing, the continuous removal of surface material significantly decreases the usable lifetimes of many devices such as automobile engines (*1*), artificial joints (*2*), and computer hard drives (*3*). The enormous economic and environmental damage due to uncontrolled friction and wear (*4*), as well as the desire to understand the relevant processes, have spurred interdisciplinary research activity

[1]Department of Chemistry, [2]Department of Applied Mathematics, University of Western Ontario, London, Ontario, Canada, N6A 5B7.

*To whom correspondence should be addressed. E-mail: mmuser@uwo.ca

with a new emphasis on the nanometer scale (5–10).

One method commonly used to protect surfaces from wear involves the use of anti-wear additives. These additives are incorporated into lubricant packages and form surface films that protect the underlying material from the destructive forces applied under sliding conditions. The most widespread application of antiwear additives is in automobile engines, where the rubbing surfaces are almost exclusively composed of steel or cast iron. The most common antiwear additives used in engines are zinc dialkyldithiophosphates (ZDDPs), which have the chemical formula $Zn[S_2P(OR)_2]_2$, where R is an alkyl group. ZDDPs have successfully been used for over 60 years, and to date no superior antiwear additive has been developed for use on steel (11).

The need for new wear inhibitors has arisen in response to environmental concerns associated with the deleterious effects that Zn, S, and P have on catalytic converters (12) and because there is substantial interest in replacing steel engines with ones composed of aluminum. The inability of ZDDP antiwear films to protect aluminum surfaces has been a major impediment to the mass production of automobiles that contain engines in which aluminum is the main component (13). There is thus a desire to design new antiwear additives. Unfortunately, the physical and chemical processes responsible for the formation and function of antiwear films are generally unknown, and as a result, current antiwear additive development strategies are largely based on trial and error instead of rational design.

Experiments clearly indicate that ZDDPs decompose in oil to form zinc polyphosphate chains that accumulate on surfaces and are transformed into a zinc phosphate film with a significantly reduced sulfur and hydrocarbon content (14, 15). Current models of their formation (14, 16) employ complex thermally activated surface reactions at low pressure and consider zinc to be a spectator ion that merely balances the charge. These theories fail to account for the extreme pressures at the intimate points of contact, nor do they explain why zinc cannot be replaced by other dications, particularly calcium (17). The rapid rate at which films form (18), however, may point to a pressure-induced bulk effect rather than a surface-mediated growth process. Moreover, it has been shown with nanometer resolution that the films on top of asperities, which experience higher pressures, are harder than those between asperities and exhibit chemical spectra suggestive of longer-chain polyphosphates (19). Despite these recent studies and decades of research on ZDDPs (20, 21), the molecular structure and hence the functionality of antiwear films have not yet been

Fig. 1. Instantaneous normalized volume $V/V_{ref} + p/p_{ref}$ as a function of pressure p for the zinc phosphate system with $V_{ref} = 200$ cm³/mol and $p_{ref} = 85$ GPa. The data for the initial c/d cycle are shown only up to 15 GPa for the sake of clarity. Beyond this point, the data obtained during the initial cycle were similar to those obtained through the second c/d cycle, which is shown in full. Key points along these plots are denoted with labels A through C2. The representation of the data as $V/V_{ref} + p/p_{ref}$ against p, instead of V against p, was performed to keep the plotted quantities in a narrow range. The bulk modulus B can be approximated in units of GPa as $B = (-\text{local slope} + 1/85)^{-1}$.
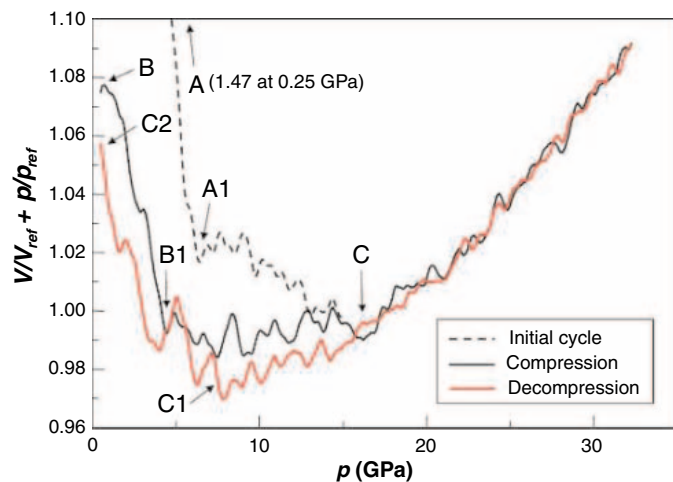
determined. Shedding light on the molecular dynamics of antiwear films would thus greatly assist the development of a molecular theory with predictive power.

Computer simulation has proven to be an effective means of studying the molecular-level details of physical processes. Although both the physics and the chemistry of sliding contacts have been modeled with ever-increasing accuracy (22–24), even allowing a direct comparison with experimental results (25), most tribological simulations have neglected the complexity of commercial lubricant systems and instead have focused on relatively simple hydrocarbon models, for which sufficiently accurate potentials are available. Conversely, quantum chemical simulations of lubricant additives have been performed (26, 27) but have failed to incorporate the extreme physical conditions that are encountered in engineering contacts.

Here we report quantum chemical simulations of tribochemical reactions under extreme nonequilibrium conditions. The results of these simulations suggest that the formation, functionality, and frictional properties of ZDDP antiwear films are due to pressure-induced changes in the bonding at Zn atoms that transform an initially visco-elastic system of zinc phosphate chains into a chemically cross-linked network. The degree of connectivity within the system, and the resulting elastic properties, are found to depend sensitively on the maximum pressure to which the system has been exposed, in a manner that correlates well with the known properties of ZDDPs on steel surfaces. Because atoms with flexible coordination numbers other than Zn can act as cross-linking agents, we suggest that this mechanism can be extended to other systems. For example, similar processes may occur in artificial joints, where it is observed that proteins decompose and form interconnected carbon-based sheets similar to graphite that prevent wear (28).

Key features in tribological systems are the flash pressures p_{max} and flash temperatures T_{max} encountered at the microscopic points of contact. p_{max} and T_{max} can be close to or even slightly exceed the theoretical yield strength p_y and melting point of the surface material (29). To investigate the effect of such extreme conditions on industrial antiwear films, we simulated the response of bulk zinc phosphate to various values of p_{max}. For this purpose, parameter-free Car-Parrinello ab initio molecular dynamics (AIMD) (30, 31) simulations were performed. AIMD simulations have recently made enormous strides in the simulation of materials under extreme conditions (32, 33). Here, AIMD simulations were performed on systems composed of one triphosphate molecule ($P_3O_{10}H_5$) and either one or two zinc phosphate molecules ($Zn[PO_4H_2]_2$) per simulation cell, using periodic boundary conditions (34). To mimic the aforementioned conditions, these systems were initially equilibrated at a pressure of 0.25 GPa. The pressure was then linearly increased at a rate of either 2.5 or 10.0 GPa/ps to p_{max} values of 2.5, 4.0, 7.0, 22.5, or 32.5 GPa. These choices for p_{max} were motivated by the theoretical yield strengths of typical aluminum (~7 GPa) and steel (~21 GPa) alloys. In all cases, once p_{max} was reached, the pressure was decreased to 0.25 GPa at the same rate. This procedure will be referred to as a compression/decompression (c/d) cycle. The results of all simulations were in general agreement, irrespective of compression rate, temperature, initial alignment of the molecules, or system size. In what follows, the data obtained from a $p_{max} = 32.5$ GPa simulation will be explicitly considered.

The calculated equation of state for the zinc phosphate system is shown in Fig. 1. At

the start of the simulation (point A in Fig. 1 and Fig. 2A), the system is in a relatively disconnected state, as would be expected after the initial ZDDP decomposition (*21*). During the initial compression, the Zn atom fluctuates between di-, tri-, and tetra-coordinate bonding arrangements. At a pressure of ~ 6 GPa (point A1 in Fig. 1), a see-saw coordination geometry (*35*) is adopted by the Zn sites. At this point, the bonding arrangement is irreversibly altered so that a cross-linked structure, as shown in Fig. 2B, will be found after decompression. The increase in the density relative to the initial zinc phosphate chains depends on p_{max}. The newly formed, cross-linked, low-pressure structure (point B in Fig. 1) serves as the starting point for a subsequent c/d cycle. The see-saw geometry at the Zn centers is recovered at a similar pressure (point B1 in Fig. 1) as in the initial cycle. When pressures exceed ~ 17 GPa (point C in Fig. 1), all simulations reveal that a highly cross-linked configuration is adopted in which the Zn is hexacoordinate. The resulting high-pressure structure exhibits cross-linking in all three directions through the Zn atom, as shown in Fig. 2C, and persists until p_{max} is reached. The system remains in the highly cross-linked state until the Zn atom returns to a tetrahedral geometry at ~ 7 GPa during decompression (point C1 in Fig. 1). Reducing the pressure to 0.25 GPa (point C2 in Fig. 1) does not result in any further modification of the internal connectivity. The final low-pressure configuration, as shown in Fig. 2B, contains tetracoordinate Zn atoms. This structure constitutes a fully connected network, unlike structure A.

Changes in the chemical connectivity alter the elastic properties of the material. The formation of highly cross-linked networks at 17 GPa increases the bulk modulus B and shear modulus of the films. The calculated value of B is 140 GPa, which is slightly less than that of iron (170 GPa). This indicates that the antiwear film will be more compliant than an underlying iron surface when exposed to high pressures. Therefore, the antiwear films do not abrade nor will they become embedded in the substrate, which are attributes necessary for effective wear inhibition. In addition, the peak pressures to which the iron surface is exposed will be reduced. The pressure-induced hardening of zinc phosphates contrasts with the widely held theory that ZDDP antiwear films primarily act passively as sacrificial coatings that undergo wear to protect the underlying surface.

The transition between states with tetra- and hexacoordinate Zn atoms occurred at different pressures during compression and decompression, resulting in the pressure-volume hysteresis between 7 and 17 GPa (points C and C1 in Fig. 1). The end points of this interval coincide with the formation and dissociation of chemical bonds; that is, with points where the previous structure becomes unstable. Instabilities generally lead to hystereses and consequently to (kinetic) friction with logarithmic-type dependencies on compression rate or sliding velocity (*36*): a characteristic of solid friction laws (*37*). Integration over the hysteresis loop between 7 and 17 GPa in Fig. 1 yielded an estimate of 18.9 kcal/mol for the dissipated energy per cycle. Therefore, as is the case for rubber (*38*), we suggest that a significant fraction of the energy loss under sliding conditions occurs in the zinc phosphate bulk.

Our simulations explain a diverse body of experimental data. Antiwear films have been referred to as "smart materials" because the harder one pushes on them, the harder they become (*39*). The current study demonstrates that this effect is due to an increase in the cross-link density with increasing pressure.

Moreover, our simulations point toward a chemical memory effect, whereby the density of a decompressed film and its bulk modulus increase with the maximum pressure to which it has been exposed. This explains why films formed on top of asperities have higher bulk and indentation moduli than those in the valleys between asperities (*40*). Our scenario also interprets the abrupt improvement in the antiwear capabilities of ZDDP films formed on steel, whose hardness was steadily increased by tempering and not through chemical modification (*41*). Namely, once p_y is sufficiently high to allow for hexacoordinate Zn, the antiwear functionality of the films is fully activated. Conversely, on materials with a value of p_y that is roughly between 5 and 17 GPa, (points A1 and C in Fig. 1), ZDDP films will form but will not effectively inhibit wear, because the pressures required for the formation of hexacoordinate Zn cannot be reached. Indeed, aluminum surfaces ($p_y \sim 7$ GPa) deteriorate significantly despite the presence of films that are similar in composition to those found on steel (*42*). Even worse, the bulk modulus of the tetra-coordinate cross-linked zinc phosphate network surpasses that of aluminum (~ 70 GPa). This explains why films have been observed to abrade and become embedded in aluminum surfaces (*42*). If p_y of the underlying substrate is less than ~ 5 GPa, then the tetracoordinate cross-links are unlikely to form altogether. Consequently, hardening due to chemical changes would not occur. This is consistent with the rather modest work hardening seen in silver polyphosphates that have been exposed to only 0.5 GPa (*43*). Finally, one can expect that the antiwear film functionality would be reduced if some of the Zn atoms were replaced by cations without flexible coordination. These cations would act as network modifiers, which in fact decrease the connectivity and hence the hardness of the film. This explains why lubricants with high concentrations of calcium phosphates lead to increased wear on steel (*17*).

In summary, computer simulations of zinc phosphates under extreme conditions indicate that pressure-induced cross-linking is a key mechanism in the formation and functionality of antiwear films. In terms of immediate application, the mechanism suggested in this study may aid in the development of ZDDP analogs for use in aluminum engines. Systems other than zinc phosphates also have the ability to form pressure-induced cross-links, and they exhibit similar modifications of elastic properties. It will be necessary to develop a material that forms a highly cross-linked structure below the p_y of aluminum, with a bulk modulus similar to but slightly lower than that of aluminum.
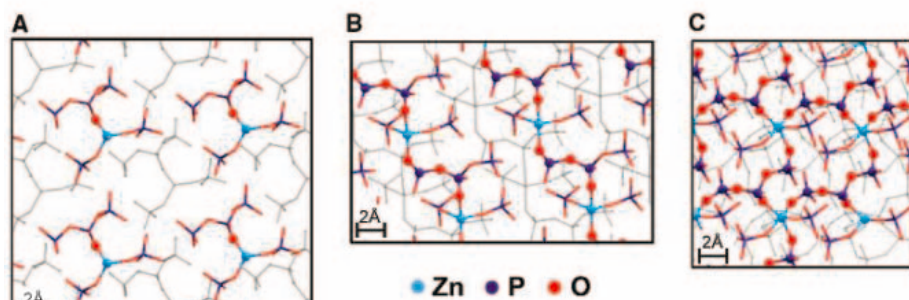
Fig. 2. Representative structures observed during the simulations of the zinc phosphate system. In all cases, a set of $2 \times 2 \times 2$ (x, y, z) unit cells is shown, and the plane of the paper is taken as the x-z plane. (**A**) The starting structure of the initial cycle. (**B**) The final structure of that cycle. (**C**) The structure of the zinc phosphate system when compressed to 17 GPa. Cross-linking in (C) occurs in all three dimensions but is only shown along the plane of the page for clarity. Color is used to indicate atoms in the uppermost layer; the atoms in the layer underneath are shown in gray. In the uppermost layer, ball-and-stick representations are used to denote connectivity within the system, whereas cylinders indicate atoms not involved in extended bonding. Hydrogen atoms have been omitted for clarity.

References and Notes

1. W. J. Bartz, Ed., *Engines and Automotive Lubrication* (Marcel Dekker, New York, 1993).
2. I. M. Hutchings, Ed., *Friction, Lubrication and Wear of Artificial Joints* (Professional Engineering Publishing, Bury St. Edmunds, UK, 2003).
3. B. Bhushan, *Tribology and Mechanics of Magnetic Storage Devices* (Springer-Verlag, New York, ed. 2, 1996).
4. H. P. Jost, *Wear* 136, 1 (1990).
5. M. Urbakh, J. Klafter, D. Gourdon, J. Israelachvili, *Nature* 430, 525 (2004).
6. U. Raviv, J. Klein, *Science* 297, 1540 (2002).
7. B. Bhushan, J. B. Israelachvili, U. Landman, *Nature* 374, 607 (1995).
8. S. Granick, *Phys. Today* 52, 26 (1999).
9. M. Cieplak, E. D. Smith, M. O. Robbins, *Science* 265, 1209 (1994).
10. M. Abdelmaksoud, J. W. Bender, J. Krim, *Phys. Rev. Lett.* 92, 176101 (2004).
11. Z. Pawlak, *Tribochemistry of Lubricating Oils* (Elsevier, Amsterdam, 2003).
12. This has led to stricter controls on the concentrations of these elements in engine oils, which has limited the amount of ZDDP that can be incorporated into motor oil formulations.
13. The replacement of steel by aluminum is motivated by efforts to reduce vehicle weight as a means of improving fuel efficiency. For example, it has been estimated that a 10% reduction in vehicle weight can result in a 7% improvement in fuel economy. However, because of the inability of ZDDPs to adequately protect aluminum surfaces, automobile manufacturers have had to resort to engines composed of aluminum-based composite materials or to engine blocks that contain steel sleeves. These measures are costly and complicate engine fabrication.
14. M. L. Suominen Fuller, M. Kasrai, G. M. Bancroft, K. Fyfe, K. H. Tan, *Tribol. Int.* 31, 627 (1999).
15. J. M. Martin, C. Grossiord, T. Le Mogne, S. Bec, A. Tonck, *Tribol. Int.* 34, 523 (2001).
16. P. A. Willermet, D. P. Dailey, R. O. Carter III, P. J. Schmitz, W. Zhu, *Tribol. Int.* 28, 177 (1995).
17. Y. Wan *et al.*, *Tribol. Ser.* 40, 155 (2002).
18. Experiments show that zinc phosphate films almost 100 nm high can be formed after a few minutes of rubbing (44). The real contact time at the molecular level is significantly less than the rubbing time, because only a small fraction of the asperities are in contact with the opposing surface at a given time.
19. M. A. Nicholls *et al.*, *Tribol. Lett.* 17, 205 (2004).
20. H. Spikes, *Tribol. Lett.* 17, 469 (2004).
21. M. A. Nicholls, T. Do, P. R. Norton, M. Kasrai, G. M. Bancroft, *Tribol. Int.* 38, 15 (2005).
22. A. B. Tutein, S. J. Stuart, J. A. Harrison, *Langmuir* 16, 291 (2000).
23. G. He, M. H. Müser, M. O. Robbins, *Science* 284, 1650 (1999).
24. J. A. Harrison, D. W. Brenner, *J. Am. Chem. Soc.* 116, 10399 (1994).
25. S. Bair, C. McCabe, P. T. Cummings, *Phys. Rev. Lett.* 88, 058302 (2002).
26. S. Jiang *et al.*, *J. Phys. Chem.* 100, 15760 (1996).
27. N. J. Mosey, T. K. Woo, *J. Phys. Chem. A* 107, 5058 (2003).
28. M. A. Wimmer, C. Sprecher, R. Hauert, G. Täger, A. Fischer, *Wear* 255, 1007 (2003).
29. U. Landman, W. D. Luedtke, J. P. Gao, *Langmuir* 12, 4514 (1996).
30. R. Car, M. Parinello, *Phys. Rev. Lett.* 55, 2471 (1985).
31. M. Parrinello, *Solid State Commun.* 38, 115 (1997).
32. C. Cavazzoni *et al.*, *Science* 283, 44 (1999).
33. D. Alfè, M. J. Gillan, G. D. Price, *Nature* 401, 462 (1999).
34. All AIMD simulations were performed with the CPMD software package (45). The potential energy was calculated using Kohn-Sham density functional theory with the gradient-corrected exchange-correlation functional of Perdew, Burke, and Ernzerhof (46); Troullier-Martins–type pseudopotentials; and a Γ-point plane wave expansion of the valence orbitals up to 120 Ry. A time step of 2.0 atomic units, equivalent to 0.0483 fs, was used in all simulations. Pressure was applied isotropically using the Parrinello-Rahman variable cell method (47, 48), and preliminary calculations showed that well-converged values for the pressure were achieved using the theoretical approach outlined above. Temperatures of 100 and 1000 K were considered, with most simulations being performed at 100 K to isolate the effect of pressure. To give the reader an idea of the computational effort associated with these simulations, it is noted that one complete c/d cycle over a range of 0.25 to 32.5 GPa at 2.5 GPa/ps required nearly 5 weeks of CPU time on a Beowulf cluster of 24 Compaq Alpha ES40 computers running at 833 MHz.
35. A four-coordinate see-saw geometry can be derived by removing any two adjacent bond sites from a central atom with an originally six-coordinate octahedral geometry.
36. M. H. Müser, *Phys. Rev. Lett.* 89, 224301 (2002).
37. The logarithmic-type dependence of the underlying chemical reactions on the compression rate can be rationalized within Eyring theory.
38. B. N. J. Persson, *J. Chem. Phys.* 115, 3840 (2001).
39. S. Bec *et al.*, *Proc. R. Soc. London Ser. A* 455, 4181 (1999).
40. J. F. Graham, C. McCague, P. R. Norton, *Tribol. Lett.* 6, 149 (1999).
41. J. S. Sheasby, T. A. Caughlin, W. A. Mackwood, *Wear* 201, 209 (1996).
42. M. Fuller *et al.*, *Tribol. Lett.* 1, 367 (1995).
43. Z. Wisniewski, R. Wisniewski, J. L. Nowinski, *Solid State Ionics* 157, 275 (2003).
44. M. Aktary, M. T. McDermott, G. A. McAlpine, *Tribol. Lett.* 12, 155 (2002).
45. J. Hütter *et al.*, *CPMD* (MPI für Festkörperforschung Stuttgart and IBM Zurich Research Laboratory, Rueschlikon, Switzerland, 1995-2001).
46. J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77, 3865 (1996).
47. M. Parrinello, A. Rahman, *Phys. Rev. Lett.* 45, 1196 (1980).
48. P. Focher, G. L. Chiarotti, M. Bernasconi, E. Tosatti, M. Parrinello, *Europhys. Lett.* 36, 345 (1994).
49. We thank P. R. Norton for inspiring us to work on this topic and P. R. Norton, M. Kasrai, Y.-T. Cheng, and W. Capehart for many useful discussions. The Natural Science and Engineering Research Council of Canada and General Motors R&D are acknowledged for providing financial support. Computational resources were made available by the Canadian Foundation for Innovation and SHARCNet of Canada.

High-Resolution Surface-Wave Tomography from Ambient Seismic Noise

Nikolai M. Shapiro,[1]* Michel Campillo,[2] Laurent Stehly,[2] Michael H. Ritzwoller[1]

Cross-correlation of 1 month of ambient seismic noise recorded at USArray stations in California yields hundreds of short-period surface-wave group-speed measurements on interstation paths. We used these measurements to construct tomographic images of the principal geological units of California, with low-speed anomalies corresponding to the main sedimentary basins and high-speed anomalies corresponding to the igneous cores of the major mountain ranges. This method can improve the resolution and fidelity of crustal images obtained from surface-wave analyses.

The aim of ambitious new deployments of seismic arrays, such as the Program for the Array Seismic Studies of the Continental Lithosphere (PASSCAL) and USArray programs (1), is to improve the resolution of images of Earth's interior by adding more instruments to regional- and continental-scale seismic networks. Traditional observational methods cannot fully exploit emerging array data because they are based on seismic waves emitted from earthquakes, which emanate from select source regions predominantly near plate boundaries and are observed at stations far from the source regions, such as most locations within the United States. With such teleseismic observations, high-frequency information is lost because of intrinsic attenuation and scattering, and resolution is degraded by the spatial extent of the surface wave's sensitivity, which expands with path length (2–4). We have moved beyond the limitations of methods based on earthquakes and recovered surface-wave dispersion data from ambient seismic noise (5).

The basic idea of the new method is that cross-correlation of a random isotropic wavefield computed between a pair of receivers will result in a waveform that differs only by an amplitude factor from the Green function between the receivers (6, 7). This property is reminiscent of the fluctuation-dissipation theorem (8), which posits a relation between the random fluctuations of a linear system and the system's response to an external force. The relation is widely used in a variety of physical applications and has its roots in early works on Brownian noise (9, 10). Recent results in helioseismology (11), acoustics (12–16), and seismology (5, 17)

[1]Center for Imaging the Earth's Interior, Department of Physics, University of Colorado at Boulder, Boulder, CO, USA. [2]Laboratoire de Géophysique Interne et de Tectonophysique, Université Joseph Fourier, Grenoble, France.

*To whom correspondence should be addressed. E-mail: nshapiro@ciei.colorado.edu
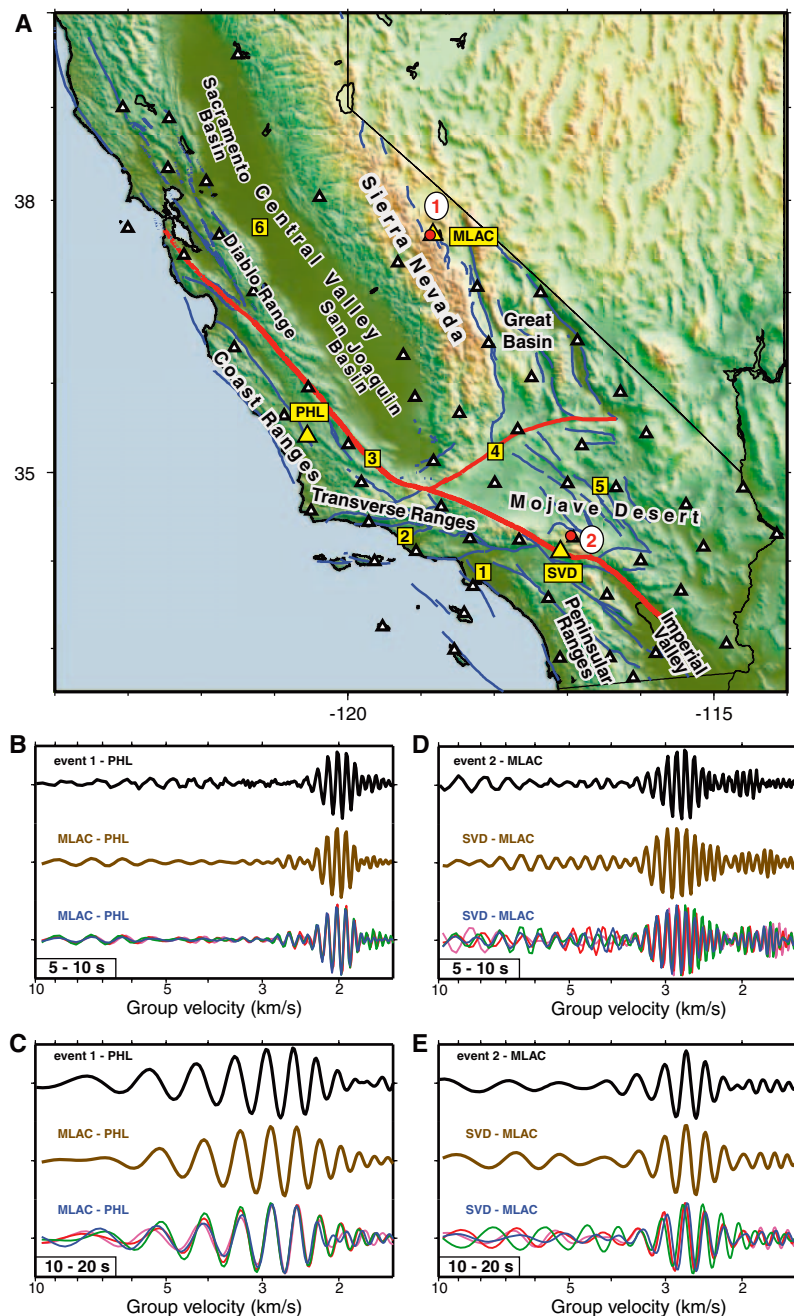
Fig. 1. Waveforms emerging from cross-correlations of ambient seismic noise compared with Rayleigh waves excited by earthquakes. (**A**) Reference map showing the locations of the principal geographical and geological features discussed in the text. White triangles show the locations of the USArray stations used in this study (5 of the 62 stations are located north of 40°N). Blue and red solid lines are the locations of known active faults. Yellow rectangles with digits indicate the following features: (1) Los Angeles Basin, (2) Ventura Basin, (3) San Andreas Fault, (4) Garlock Fault, (5) Mojave shear zone, and (6) Stockton Arch. (**B**) Comparison of waves propagating between stations MLAC and PHL [yellow triangles in (A)], bandpassed over periods between 5 and 10 s. The upper trace (black) is the signal emitted by earthquake 1 [white circle with red number in (A)] near MLAC observed at PHL; the middle trace (gold) is the cross-correlation from 1 year of ambient seismic noise observed at stations MLAC and PHL; and the lower traces are cross-correlations from 4 separate months of noise observed at the two stations in 2002 (magenta, January; red, April; green, July; blue, October). The earthquake-emitted signal was normalized to the spectrum of the cross-correlated ambient noise. (**C**) Similar to (B), but with the bandpass filter at periods between 10 and 20 s. (**D**) Similar to (B), but between stations SVD and MLAC [yellow triangles in (A)]. Earthquake 2 is near station SVD, observed at station MLAC. (**E**) Similar to (D), but with the bandpass filter at periods between 10 and 20 s.

suggest that such a statistical treatment can be applied to nonthermal random wavefields, in particular to long series of ambient seismic noise, because the distribution of the ambient sources randomizes when averaged over long times. Ambient seismic noise is additionally randomized by scattering from heterogeneities within Earth (*18*). Surface waves are most easily extracted from the ambient noise (*5*), because they dominate the Green function between receivers located at the surface and also because ambient seismic noise is excited preferentially by superficial sources, such as oceanic microseisms and atmospheric disturbances (*19–22*). The seismic noise field is often not perfectly isotropic and may be dominated by waves arriving from a few principal directions. To reduce the contribution of the most energetic arrivals, we disregard the amplitude by correlating only one-bit signals (*15, 17*) before the computation of the cross-correlation.

Examples of cross-correlations between pairs of seismic stations in California appear in Fig. 1 (*23*). Cross-correlations between two station pairs (MLAC-PHL and SVD-MLAC) in two short-period bands (5 to 10 s and 10 to 20 s) are presented using four different 1-month time series (January, April, July, and October 2002). For each station pair, results from different months are similar to one another and to the results produced by analyzing a whole year of data, but differ between the station pairs. Thus, the emerging waveforms are stable over time and characterize the structure of the earth between the stations. In addition, the cross-correlations of noise sequences are very similar to surface waves emitted by earthquakes near one receiver observed at the other receiver. This confirms that the cross-correlations approximate Green functions of Rayleigh waves propagating between each pair of stations and that 1 month of data suffices to extract Rayleigh-wave Green functions robustly in the period band of interest here (7 to 20 s).

We selected 30 relatively quiescent days (during which no earthquakes stronger than magnitude 5.8 occurred) of continuous data taken at a rate of one sample per second from 62 USArray stations within California (*24*) during August and September 2004. Short-period surface-wave dispersion curves are estimated from the Green functions using frequency-time analysis (*25–27*) from the 1891 paths connecting these stations. We rejected waveforms with signal-to-noise ratios smaller than 4 and for paths shorter than two wavelengths, resulting in 678 and 891 group-speed measurements at periods of 7.5 and 15 s, respectively (fig. S2). We then applied a tomographic inversion (*28*) to these two data sets to obtain group-speed maps on a 28 × 28 km grid across California (Fig. 2).
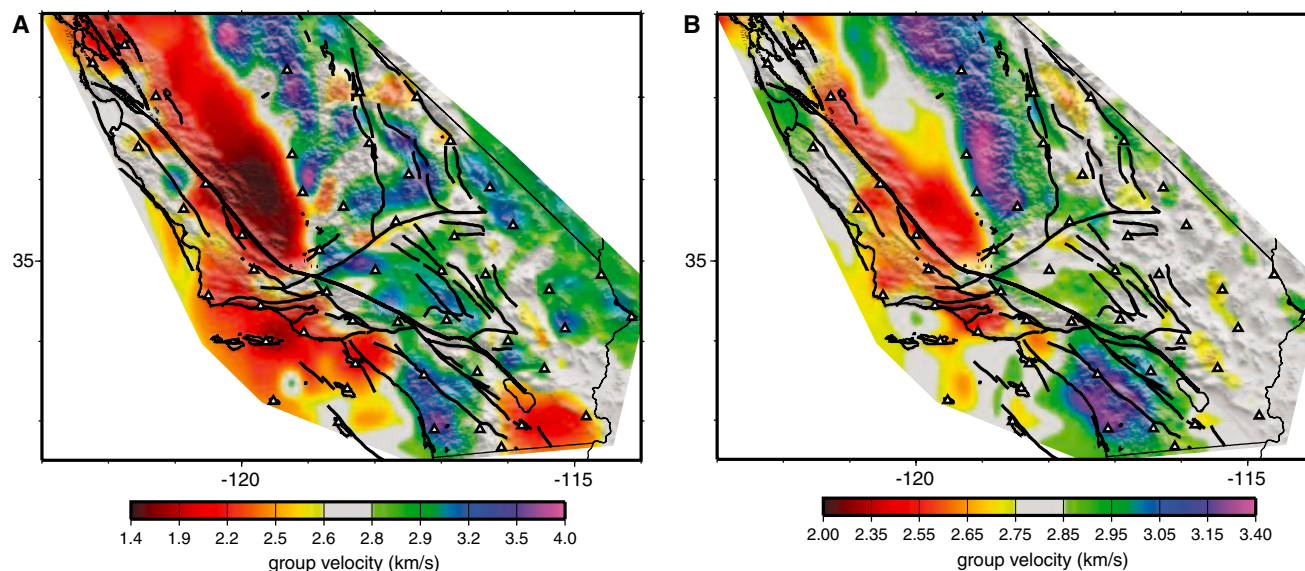
Fig. 2. Group-speed maps constructed by cross-correlating 30 days of ambient noise between USArray stations. (**A**) 7.5-s-period Rayleigh waves. (**B**) 15-s-period Rayleigh waves. Black solid lines show known active faults. White triangles show locations of USArray stations used in this study. Similar maps from a different single month of data are shown in the supporting online material.

The maps produced variance reductions of 93 and 76% at 7.5 and 15 s, respectively, relative to the regional average speed at each period. To test the robustness of the inversion, we applied the same procedure to a second month of data and produced similar tomographic maps (fig. S3). The resolution of the resulting images is about the average interstation distance, between 60 and 100 km across most of each map (fig. S4).

A variety of geological features (*29*) are recognizable in the estimated group-speed dispersion maps (Fig. 2). For the 7.5-s Rayleigh wave, which is most sensitive to shallow crustal structures no deeper than about 10 km, the dispersion map displays low group speeds for the principal sedimentary basins in California, including the basins in the Central Valley, the Salton Trough in the Imperial Valley, the Los Angeles Basin, and the Ventura Basin. Regions consisting mainly of plutonic rocks (the Sierra Nevada, the Peninsular Ranges, the Great Basin, and the Mojave Desert region) are characterized predominantly by fast group speeds. Somewhat lower speeds are observed in the Mojave Shear Zone and along the Garlock Fault. The Coast Ranges, the Transverse Ranges, and the Diablo Range, which are mainly composed of sedimentary rocks, are characterized by low group speeds, with the exception of the Salinian block located south of Monterey Bay.

For the 15-s Rayleigh wave, which is sensitive mainly to the middle crust down to depths of about 20 km, very fast group speeds correspond to the remnants of the Mesozoic volcanic arc: the Sierra Nevada and the Peninsular Ranges, composed principally of Cretaceous granitic batholiths. The

map also reveals the contrast between the western and eastern parts of the Sierra Nevada (*30*). The group speeds are lower in the Great Basin and in the Mojave Desert, indicating that the middle crust in these areas is probably hotter and weaker than in the Sierra Nevada. In the Central Valley, slow group speeds are associated with two deep sedimentary basins: the San Joaquin Basin in the south and the Sacramento Basin in the north, separated in the middle by the igneous-dominated Stockton Arch (*31*). Group speeds are low in the sedimentary mountain ranges (the Transverse Ranges, the southern part of the Coast Ranges, and the Diablo Range). Neutral to fast wave speeds are observed for the Salinian block. In this area, the 15-s map shows a contrast between the high-speed western wall of the San Andreas Fault, composed of plutonic rocks of the Salinian block, and its low-speed eastern wall, composed of sedimentary rocks of the Franciscan formation.

These results establish that Rayleigh-wave Green functions extracted by cross-correlating long sequences of ambient seismic noise, which are discarded as part of traditional seismic data processing, contain information about the structure of the shallow and middle crust. The use of ambient seismic noise as the source of seismic observations addresses several shortcomings of traditional surface-wave methods. The method is particularly advantageous in the context of temporary seismic arrays such as the Transportable Array component of USArray or PASSCAL experiments, because it can return useful information even if earthquakes do not occur. The short-period dispersion maps produced by the method can

provide homogeneously distributed information about shear wave speeds in the crust, which are hard to acquire with traditional methods. The new method enhances resolution because measurements are made between regularly spaced receivers, which may lie much closer to one another than to earthquakes.

It may seem initially surprising that deterministic information about Earth's crust can result from correlations of ambient seismic noise. This result reminds us that random fluctuations can, in fact, yield the same information as that provided by probing a system with an external force (*9*) and that not all noise is bad. In seismology, external probing through active seismic sources (such as explosions) may be prohibitively expensive, and earthquakes are both infrequent and inhomogeneously distributed. In many instances, merely "listening" to ambient noise may be a more reliable and economical alternative.

References and Notes
1. USArray (www.iris.iris.edu/USArray) is one of the components of the new EarthScope (www.earthscope.org) initiative in the United States. PASSCAL (www.iris.edu/about/PASSCAL) is a program of the Incorporated Research Institutions for Seismology (IRIS) (www.iris.edu).
2. G. Nolet, F. A. Dahlen, *J. Geophys. Res.* **105**, 19043 (2000).
3. J. Spetzler, J. Trampert, R. Snieder, *Geophys. J. Int.* **149**, 755 (2002).
4. M. H. Ritzwoller, N. M. Shapiro, M. P. Barmin, A. L. Levshin, *J. Geophys. Res.* **107**, 2235 (2002).
5. N. M. Shapiro, M. Campillo, *Geophys. Res. Lett.* **31**, L07614, 10.1029/2004GL019491 (2004).
6. R. L. Weaver, O. I. Lobkis, *Phys. Rev. Lett.* **87**, paper 134301 (2001).
7. R. Snieder, *Phys. Rev. E* **69**, 046610 (2004).
8. R. Kubo, *Rep. Prog. Phys.* **29**, 255 (1966).
9. S. Kos, P. Littlewood, *Nature* **431**, 29 (2004).
10. A. Einstein, *Ann. Phys.* **17**, 549 (1905).

11. T. L. Duvall, S. M. Jefferies, J. W. Harvey, M. A. Pomerantz, *Nature* **362**, 430 (1993).
12. R. L. Weaver, O. I. Lobkis, *J. Acoust. Soc. Am.* **110**, 3011 (2001).
13. A. Derode *et al.*, *J. Acoust. Soc. Am.* **113**, 2973 (2003).
14. P. Roux, W. A. Kuperman, *J. Acoust. Soc. Am.* **116**, 1995 (2004).
15. E. Larose, A. Derode, M. Campillo, M. Fink, *J. Appl. Phys.* **95**, 8393 (2004).
16. A. E. Malcolm, J. A. Scales, B. A. van Tiggelen, *Phys. Rev. E* **70**, 10.1103/PhysRevE.70.015601 (2004).
17. M. Campillo, A. Paul, *Science* **299**, 547 (2003).
18. R. Hennino *et al.*, *Phys. Rev. Lett.* **86**, 3447 (2001).
19. A. Friedrich, F. Kruger, K. Klinge, *J. Seismol.* **2**, 47 (1998).
20. T. Tanimoto, *Geophys. J. Int.* **136**, 395 (1999).
21. G. Ekström, *J. Geophys. Res.* **106**, 26483 (2001).
22. J. Rhie, B. Romanowicz, *Nature* **431**, 552 (2004).
23. Data processing was performed with the Seismic Analysis Code (SAC) (*32*).
24. We used 62 stations of the Transportable Array component of USArray in California (Fig. 1A). This includes 40 permanent stations of the Southern California TriNet system (www.trinet.org), 17 permanent stations of the Berkeley Digital Seismic Network (quake.geo.berkeley.edu/bdsn), 2 permanent stations of the Anza Seismic Network (eqinfo.ucsd.edu/deployments/anza.html), and 3 new USArray stations.
25. A. L. Levshin *et al.*, *Seismic Surface Waves in a Laterally Inhomogeneous Earth*, V. I. Keilis-Borok, Ed. (Kluwer Academic, Norwell, MA, 1989).
26. M. H. Ritzwoller, A. L. Levshin, *J. Geophys. Res.* **103**, 4839 (1998).
27. N. M. Shapiro, S. K. Singh, *Bull. Seism. Soc. Am.* **89**, 1138 (1999).
28. M. P. Barmin, M. H. Ritzwoller, A. L. Levshin, *Pure Appl. Geophys.* **158**, 1351 (2001).
29. C. W. Jennings, geologic map of California (California Division of Mines and Geology, Map No. 2, 1977).
30. C. H. Jones, H. Kanamori, S. W. Roecker, *J. Geophys. Res.* **99**, 4567 (1994).
31. A. M. Wilson, G. Garven, J. R. Boles, *GSA Bull.* **111**, 432 (1999).
32. P. Goldstein, L. Minner, *Seism. Res. Lett.* **67**, 39 (1996) (www.llnl.gov/sac).
33. The data used in this work were obtained from the IRIS Data Management Center. We are also particularly grateful to M. Barmin for help with the tomographic code and P. Goldstein for clarifications about the SAC program. We thank C. Jones for a tutorial on the geology of California and E. Larose, O. Lobkis, L. Margerin, R. Maynard, A. Paul, B. van Tiggelen, and R. Weaver for helpful discussions. We acknowledge the support from CNRS/Institut National des Sciences de l'Univers (program DyETI) and the Commissariat à l'Energie Atomique (France).

Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication

Greger Larson,[1]* Keith Dobney,[2] Umberto Albarella,[3]
Meiying Fang,[4] Elizabeth Matisoo-Smith,[5] Judith Robins,[5]
Stewart Lowden,[6] Heather Finlayson,[7] Tina Brand,[8]
Eske Willerslev,[1] Peter Rowley-Conwy,[2]
Leif Andersson,[4] Alan Cooper[1]*†

Mitochondrial DNA (mtDNA) sequences from 686 wild and domestic pig specimens place the origin of wild boar in island Southeast Asia (ISEA), where they dispersed across Eurasia. Previous morphological and genetic evidence suggested pig domestication took place in a limited number of locations (principally the Near East and Far East). In contrast, new genetic data reveal multiple centers of domestication across Eurasia and that European, rather than Near Eastern, wild boar are the principal source of modern European domestic pigs.

The domestication of plants and animals led to one of the most important socioeconomic transitions in human history, yet little is known about whether the process took place in a limited number of geographic regions or was a more widespread innovation involving multiple, independent "events." Wild boar were important prey animals for early hunter-gatherers across wide areas of Eurasia (*1*) until the early Holocene, when this predator-prey relation radically shifted as they, and several other large mammals, were domesticated. An extensive zooarcheological record suggests that pigs were first domesticated ~9000 years ago in the Near East (*2*), whereas more recent molecular and archaeological evidence suggests a second, independent domestication in the Far East (*3, 4*). In eastern Anatolia, several sites record gradual changes in pig morphology and demographic profile (principally a reduction in certain tooth dimensions and the increased predominance of younger animals in archaeological assemblages) (*5, 6*) over several millennia, and these have been taken to represent the domestication process in situ. Although the independent domestication of wild boar in Europe has been suggested (*7*), others have concluded that, like cattle (*8*) and sheep, pigs derived from Near Eastern genetic stock were imported by Neolithic farmers into Europe (*9*).

The wild progenitors of many Eurasian domesticates are either extinct [e.g., the aurochs (*8*) and the wild horse (*10*)] or have little or no phylogeographic structure [e.g., the wolf (*11*)]. Consequently, the broad distribution of surviving wild boar populations across the Old World provides a unique opportunity to analyze the origins of modern domestic lineages. Previous studies (*3, 12*) have identified three divergent clusters of *Sus scrofa* mitochondrial sequences, one Asian clade and two European groups, of which one consists solely of Italian wild boar. Both the Asian and European groups contain domestic breeds, yet molecular clock estimates indicate the split between the two groups significantly predates evidence for pig domestication, which suggests independent domestication events in each area from divergent wild boar lineages (*3, 12*).

To investigate the relationships between domestic pigs and indigenous wild boar across their range, we sequenced 663 base pairs (bp) of the mitochondrial control region from 165 wild and feral pigs primarily from museum specimens, using appropriate ancient-DNA methods (*13*), and from 58 domestic pigs. An additional 463 individual pig sequences were obtained from GenBank, and phylogenetic analyses were performed using Bayesian Monte Carlo–Markov chain (MCMC) (*14*) and median-joining networks (*15*). The consensus tree (Fig. 1) shows that the basal lineages of *S. scrofa* occur in western island Southeast Asia (ISEA). An initial dispersal from this area into the Indian subcontinent was followed by subsequent radiations into East Asia and a final, progressive spread across Eurasia into Western Europe. The marked East-West split among wild boar is consistent with morpho-

[1]Henry Wellcome Ancient Biomolecules Centre, University of Oxford, Department of Zoology, South Parks Road OX1 3PS, UK. [2]Department of Archaeology, University of Durham, South Road, Durham DH1 3L, UK. [3]Department of Archaeology, University of Sheffield, West Street, Sheffield S1 4ET, UK. [4]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences and Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala Biomedical Center, Box 597, SE-75124 Uppsala, Sweden. [5]Department of Anthropology and Allan Wilson Centre for Molecular Ecology and Evolution, University of Auckland, P.B. 92019, Auckland, New Zealand. [6]Jurox Pty Limited, 85 Gardiners Road, Rutherford, NSW, 2320, Australia. [7]Department of Genomics and Bioinformatics, Roslin Institute, Roslin, Midlothian, EH25 9PS, UK. [8]Department of Evolutionary Biology, Zoological Institute, University of Copenhagen, Universitetsparken 15, DK-2100, Copenhagen O, Denmark.

*To whom correspondence should be addressed. E-mail: greger.larson@zoo.ox.ac.uk (G.L.) and alan.cooper@adelaide.edu.au (A.C.)
†Present address: School of Environmental Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

logically based studies, which have highlighted the distinctiveness of the Southeast Asian forms (*Sus scrofa vittatus*) in particular (*16*). Interestingly, fewer than 2% of wild boar specimens (*13*) violate the otherwise consistent correlation between phylogeny and geography, and these are likely the result of past, human-mediated introgression and/or dispersal. This strong phylogeographic structure provides an opportunity to trace the geographic origins of domestic pig lineages.

Modern domestic European pig breeds fall within the general European cluster (excluding recently "improved" breeds known to have been crossbred with Asian varieties over the past 200 years) and lack any affinity with wild boar lineages present in the Middle East

(Fig. 1). The complete lack of Turkish, Armenian, or Iranian signatures within modern European breeds implies that, even if domesticated Near Eastern pigs entered central Europe with early Near Eastern farmers, those lineages have left no descendants among modern European domestic pigs. Network analysis (Fig. 2) provides further resolution of the closely related European haplotypes and reveals two core lineages (A+C), separated by a transversion, each of which is surrounded by a starlike pattern consistent with a recent population expansion, analogous to that seen in cattle (*8*). Statistical analyses (Fu and Li's *F* and Tajima's *D*) were consistent with a population expansion for both wild and domestic pigs in Europe and Asia, although statistical significance was

only reached for Fu and Li's *F* in the domestic populations (*13*). Unlike cattle, however, the two core pig haplotypes are only present in Europe (or European-derived populations), which indicates the independent domestication of at least two European wild boar lineages, although the intermingling of wild boar and domestic pig sequences throughout the network suggests that numerous additional wild boar lineages may also have been domesticated. The only region in which both core haplotypes are indigenous is Germany, which suggests that central Europe may have been a center for early European pig domestication. It is noteworthy that only two of the haplotypes on the network are found within both wild boar and domestic pigs; this indicates that the wild boar sampled

in this study are unlikely to be derived from feral animals.

Network analysis of Asian domestic and wild haplotypes (Fig. 3) shows starlike rapid expansion patterns similar to those in Europe but less marked. Two haplotypes (GX and AE) found in wild boar (morphologically wild individuals from the Chinese provinces of Gansu and Hunan) are shared by several East Asian domestic pigs, although the diffuse nature of the network is also compatible with domestication of more than just these two wild boar lineages. Unlike the rather ambiguous zooarcheological record for China, where claims for early pig domestication have recently been questioned (*4*), these data clearly support recent genetic studies that have identified at least one pig domestication event in East Asia (*12*). Haplotypes I, FH, and L (Fig. 3) include feral and domestic pigs found across Asia, Australia, New Zealand, Hawaii, the United States, and Europe and appear to have been the primary haplotypes that were distributed worldwide. It is interesting that a genetically distinctive subclade of central Chinese wild boar, indigenous to Sichuan, Shanxi, and Shensi provinces, has not contributed to any of the modern Asian domestic breeds sampled (Fig. 1).

The presence of an Indian domestic pig sequence within the cluster of phylogeographically differentiated wild boar lineages from India (D3 in Fig. 1) indicates the existence of yet another independent domestication event and supports claims for pig domestication in India (*16*). Numerous introduced Australian pigs from the west coast of Cape York have a haplotype that falls within a cluster of differentiated wild boar from Burma and Thailand (D5 in Fig. 1). The presence of an exotic nematode and a Melanesian tick on the pigs also suggest that some pigs in Cape York are of Southeast Asian stock (*17*). The absence of any historical evidence for the introduction of Southeast Asian wild boar to Northern Australia, combined with the coat color variation in these feral pig populations, strongly supports their domestic origin and, therefore, identifies this part of Southeast Asia as yet another region where independent pig domestication has occurred.

The strong phylogeographic structure of pig mitochondrial sequences also provides clear evidence for human-mediated dispersal of this species. For example, *S. scrofa* was not present on the Mediterranean islands of Corsica and Sardinia before the seventh millennium B.C. (*18*). Thus, the extant populations of so-called miniature "wild boar" found on these islands has long been interpreted as a human-aided, prehistoric introduction of wild or domesticated (now feral) pigs. Indeed, all but 2 of the 15 wild boar sampled on both islands appear near the center of the European haplotype network (fig. S2). Two anomalous sequences on Sardinia, however, share a close affinity to the phylogenetically distinctive Italian wild boar clade found in Maremma (D4 in Fig. 1), the only region where native Italian wild boar populations have not been replaced by more recent introductions (*19*). Although the European lineages on Sardinia may represent recent introductions and/or introgressions, the presence of the distinctive Italian signature in at least two Sardinian pigs supports the idea of an early Italian origin. If the ancestors of these two pigs were originally introduced as domestic animals (that have since gone feral), these individuals must represent a lineage of pigs independently domesticated on the Italian mainland.

Our data also have implications for conflicting hypotheses regarding Pacific prehistory. Although the dates for the first human settlements in Near Oceania (the region from New Guinea to the Solomon Islands) range between 30,000 and 40,000 years ago, undisputed evidence of domesticated animals (pigs, dogs, and chickens) does not appear in the region until about 3500 to 3300 years before the present (B.P.), in association with the appearance of the Lapita cultural complex (*20*). Claims of pre-Lapita pigs as old as 10,000 B.P. have been made (*21*), and although pigs were probably present by 6000 B.P. (*22*), all of these dates remain contentious and have yet to be validated through carbon dating (*20*). The New Guinea pig haplotypes in this study cluster with pigs from Hawaii, Vanuatu, and Halmahera in a monophyletic group (here termed the "Pacific

Fig. 2. A parsimonious median-joining network depicting the relationships between European domestic breeds and European wild boar located solely in the red portion of Fig. 1 labeled D1. Node sizes are proportional to haplotype frequencies in the dataset. Colors within the nodes: yellow, domestic; green, wild; orange, feral; blue, unknown status; and red, inferred intermediate haplotypes not represented by any sampled pigs. Lines between nodes represent a single nucleotide change, except where perpendicular hashes represent single changes.
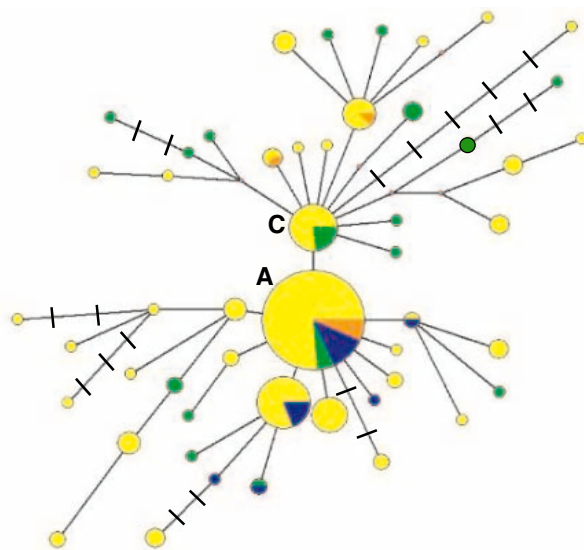
Fig. 3. A median-joining network depicting the relationships between Asian domestic breeds and closely related Asian wild boar located solely in D3 (Fig. 1). Colors are identical to Fig. 2, except gray pie segments denote samples identified as wild boar but more likely feral pigs (*13*). Only two haplotypes (AE and GX) contain both Asian domestic breeds and definitive Chinese wild boar (from Gansu and Hunan provinces, China). Japanese wild boar clusters are shown within the three shaded regions. Haplotypes I, FH, and L collectively contain feral and domestic pigs found in Asia, Australia, New Zealand, Hawaii, the United States, and Europe.
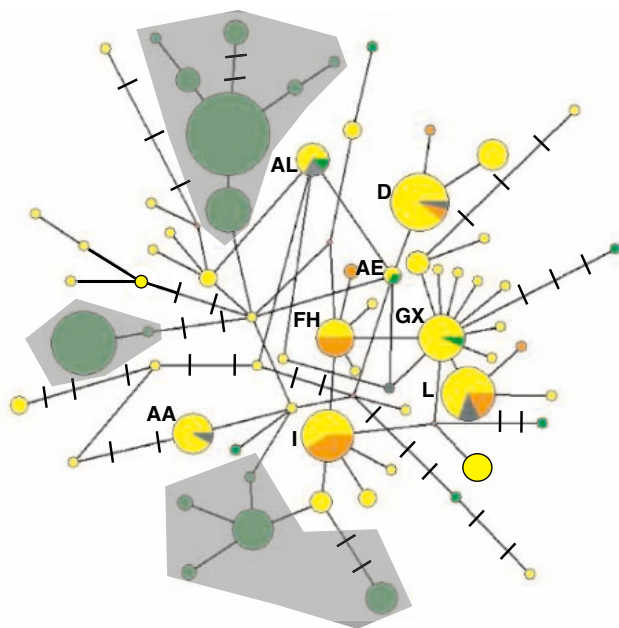
clade") within the large Eastern Eurasian cluster and are well separated from any other individuals, domestic or wild (D6 in Fig. 1). This evidence is consistent with a Lapita dispersal from Near to Remote Oceania, but the lack of any genetic affinity between this group and Taiwanese wild boar (Fig. 1) offers no support for the "Out-of-Taiwan" model (*23*) of human and pig dispersal into Near Oceania. This evidence also supports the importance of Halmahera, which has been shown to be the origin of Remote Oceanic populations of the Pacific rat (*Rattus exulans*) transported by Lapita peoples (*24*) and the origin of the human mitochondrial DNA (mtDNA) marker known as the "Polynesian motif" (*25*). The additional lack of affinity of the New Guinea pigs with *Sus celebensis* (the indigenous Sulawesi wild boar, also found on offshore islands including Halmahera) rules out a significant *S. celebensis* maternal genetic input as suggested by Groves (*16*).

The lack of an obvious genetic source population on either mainland or island Southeast Asia from which pigs in the Pacific clade were drawn is intriguing and may suggest either the existence of indigenous *S. scrofa* in Wallacea or an early human-mediated introduction from elsewhere in ISEA currently not sampled by our study. In either case, then ISEA must be considered another independent center of pig domestication. Interestingly, the two clades of *S. celebensis* (Fig. 1 and fig. S1) demonstrate that this group is not monophyletic. In fact, the North-South geographic partitioning of the two strongly implies two independent invasions of wild boar onto Sulawesi, a pattern identical to that reported for other Sulawesi fauna including macaques (*26*), shrews (*27*), and bovids (*28*). Lastly, an ISEA origin of the other sister taxa of *S. scrofa* (Javan warty pig, *S. verrucosus*, and the bearded pig, *S. barbatus*) is supported by the phylogenetic tree, although the paraphyletic arrangement provides no support for current species designations. Thus, further studies using both mtDNA and nuclear DNA are required to resolve the phylogenetic status of these taxa.

The genetic evidence presented in this study provides clear proof for multiple centers of domestication across Eurasia. To further examine the domestication of pigs, the zooarcheological records of Europe, India, Southeast Asia, and ISEA should be explored in more detail.

References and Notes

1. N. Benecke, in *Exploitation des Animaux Sauvages à travers le Temps*, J. Desse, F. Audoin-Rouzeau, Eds. (Editions APDCA, Juan-les-Pins, 1993), pp. 233–245.
2. J. Epstein, M. Bichard, in *Evolution of Domesticated Animals*, I. L. Mason, Ed. (Longman, New York, 1984), pp. 145–162.
3. E. Giuffra *et al.*, *Genetics* **154**, 1785 (2000).
4. Y. Jing, R. K. Flad, *Antiquity* **76**, 724 (2002).
5. J. Peters, D. Helmer, A. von den Driesch, M. Saña Seguí, *Paléorient* **25**, 27 (1999).
6. A. Ervynck, K. Dobney, H. Hongo, R. Meadow, *Paléorient* **27**, 47 (2002).
7. S. Bökönyi, *History of Domestic Mammals in Central and Eastern Europe* (Akademiai Kiado, Budapest, 1974).
8. C. S. Troy *et al.*, *Nature* **410**, 1088 (2001).
9. P. Rowley-Conwy, in *The Widening Harvest*, A. Ammerman, P. Biagi, Eds. (Archaeological Institute of America, Boston, 2003), pp. 99–117.
10. T. Jansen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10905 (2002).
11. P. Savolainen, Y. P. Zhang, J. Luo, J. Lundeberg, T. Leitner, *Science* **298**, 1610 (2002).
12. J. M. H. Kijas, L. Andersson, *J. Mol. Evol.* **52**, 302 (2001).
13. Materials and methods are available as supporting material on *Science* Online.
14. F. Ronquist, J. P. Huelsenbeck, *Bioinformatics* **19**, 1572 (2003).
15. H. J. Bandelt, P. Forster, A. Rohl, *Mol. Biol. Evol.* **16**, 37 (1999).
16. C. P. Groves, *Ancestors for the Pigs: Taxonomy and Phylogeny of the Genus Sus* (Tech. Bull. 3, Department of Prehistory, Research School of Pacific Studies, Australian National Univ., Canberra, 1981), 96 pp.
17. P. M. Heise-Pavlov, S. R. Heise-Pavlov, *Wildl. Biol.* **9**, 21 (2003).
18. J. D. Vigne, *Les mamméres post-glaciaires de Corse: Etude archéozoologique* (Gallia préhistoire, suppl. 26, Editions du CNRS, Paris, 1988).
19. M. Apollonio, E. Randi, S. Toso, *Boll. Zool.* **3**, 213 (1988).
20. P. V. Kirch, *On the Road of the Winds: An Archaeological History of the Pacific Islands before European Contact* (Univ. of California Press, Berkeley, CA, 2000), 424 pp.
21. S. Bulmer, *J. Soc. Océan.* **31**, 7 (1975).
22. J. Golson, P. J. Hughes, *J. Soc. Océan.* **36**, 294 (1976).
23. M. E. Hurles, E. Matisoo-Smith, R. D. Gray, D. Penny, *Trends Ecol. Evol.* **18**, 531 (2003).
24. E. Matisoo-Smith, J. H. Robins, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9167 (2004).
25. S. J. Oppenheimer, M. B. Richards, *Nature* **410**, 166 (2001).
26. B. J. Evans, J. C. Morales, J. Supriatna, D. J. Melnick, *Biol. J. Linn. Soc.* **66**, 539 (1999).
27. M. Ruedi, *Biol. J. Linn. Soc.* **58**, 197 (1996).
28. H. Kako, T. Namikawa, O. Takenaka, A. Takenaka, T. Amano, H. Martojo, *Z. Zool. Syst. Evol.* **32**, 1 (1994).
29. We thank the Wellcome and Leverhulme Trusts, the Arts and Humanities Research Board, and the European Union PigBioDiv2 project for financial support. We are especially grateful to the Smithsonian Institution Museum of Natural History and to the many other institutions and individuals listed in the SOM that provided sample material and access to collections.

High-Throughput Mapping of a Dynamic Signaling Network in Mammalian Cells

Miriam Barrios-Rodiles,[1] Kevin R. Brown,[2] Barish Ozdamar,[1,3] Rohit Bose,[1,3] Zhong Liu,[1] Robert S. Donovan,[1] Fukiko Shinjo,[1] Yongmei Liu,[1] Joanna Dembowy,[1,3] Ian W. Taylor,[1,3] Valbona Luga,[1,3] Natasa Przulj,[4] Mark Robinson,[5] Harukazu Suzuki,[6] Yoshihide Hayashizaki,[6] Igor Jurisica,[2,4,7] Jeffrey L. Wrana[1,3]*

Signaling pathways transmit information through protein interaction networks that are dynamically regulated by complex extracellular cues. We developed LUMIER (for luminescence-based mammalian interactome mapping), an automated high-throughput technology, to map protein-protein interaction networks systematically in mammalian cells and applied it to the transforming growth factor–β (TGFβ) pathway. Analysis using self-organizing maps and *k*-means clustering identified links of the TGFβ pathway to the p21-activated kinase (PAK) network, to the polarity complex, and to Occludin, a structural component of tight junctions. We show that Occludin regulates TGFβ type I receptor localization for efficient TGFβ-dependent dissolution of tight junctions during epithelial-to-mesenchymal transitions.

Dynamic protein-protein interactions (PPIs) are key for cell signaling and dictate the timing and intensity of network outputs. Systematic mapping of PPI networks has thus far focused on static analyses in *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (*1–5*). To begin building an understanding of how signaling networks convey information in vertebrates, we developed high-throughput LUMIER to systematically map PPIs in mammalian cells. This strategy uses *Renilla* luciferase enzyme (RL) fused to proteins of interest, which are then coexpressed with individual Flag-tagged partners in mammalian cells. Their interactions were determined by performing an RL enzymatic assay on immunoprecipitates using an antibody against Flag (Fig. 1A). As a model for a systematic study of cell signaling, we focused on the
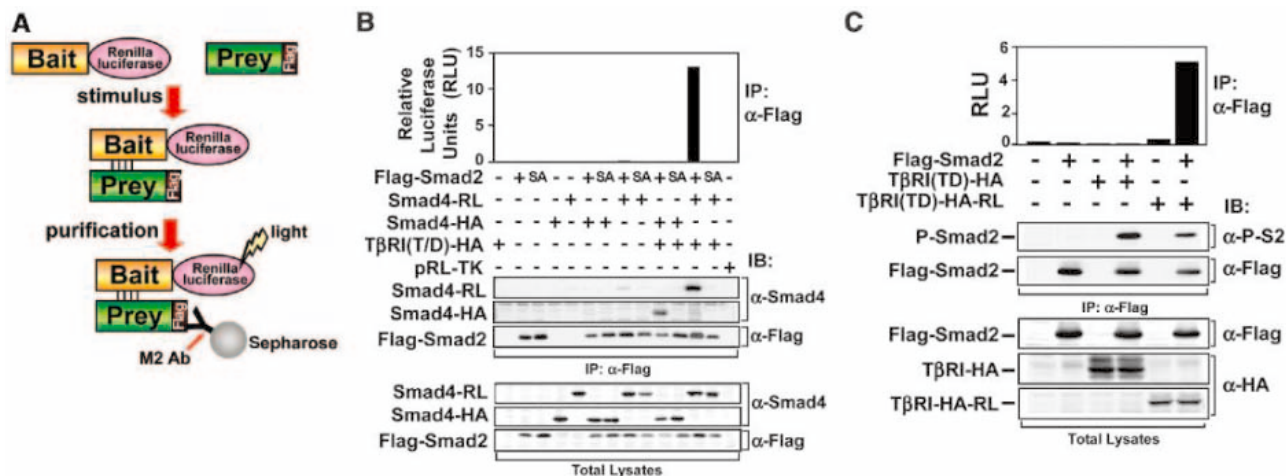
Fig. 1. A luminescence-based strategy for the detection of mammalian protein-protein interactions. (**A**) LUMIER. RL-tagged bait coexpressed with a Flag-tagged prey is detected in immunoprecipitates enzymatically as light emission. (**B**) LUMIER detects phosphorylation-dependent interactions. HEK-293T cells were transfected with hemagglutinin (HA)- or RL-tagged Smad4 together with either WT (+) or the phosphorylation site mutant (SA) of Flag-Smad2 in the absence (–) or presence (+) of TGFβ signaling. Smad4 interaction with Smad2 was determined by measuring RL activity on immunoprecipitates prepared using an antibody against Flag (histogram) or by immunoblotting (IB) with antibody against Smad4. pRL-TK is RL driven by the thymidine kinase promoter and is a negative control. (**C**) Detection of TβRI-Smad2 interaction by LUMIER. HEK-293T cells were transfected with constitutively active TβRI(T204D)-HA-RL or TβRI(T204D)-HA mutants in which Thr204 was replaced by Asp along with Flag-Smad2, and their association was detected as in (B, top panel). Levels of phosphorylated Smad2, total Smad2, and receptors were confirmed by immunoblotting with antibodies against phospho-Smad2 (α-P-S2), Flag, and HA, respectively (lower panels). Note that the TβRI(T204D)-Smad2 association cannot be detected by immunoblotting as it is a transient interaction.

Fig. 2. High-throughput LUMIER. (**A**) Pilot screen in 96-well plates. HEK-293T cells were transfected with Smad4-RL and either empty vector (NC), HA-tagged TβRI (TβRI-HA), or 30 different Flag-tagged cDNAs (indicated) in the absence (white) or presence (black) of TGFβ signaling. Their interactions were assessed by LUMIER on immunoprecipitates prepared using an antibody against Flag. Results are plotted as the mean relative luciferase activity (RLU) ± SD of triplicates with corresponding luminescence intensity ratio (LIR) values shown on the right vertical axis. (**B**) The TGFβ pathway LUMIER screen. TGFβ pathway components (listed on the left) fused to RL were screened against 518 3Flag-tagged preys (numbered at the top of the panels and grouped by domain composition as indicated on the bottom) in the presence (*) or absence of TGFβ signal. Each row corresponds to the indicated pathway component and the LIR score for each test is represented in yellow, with the intensity representing the LIR value according to the scale on the right. W, wild type; K, kinase deficient; Q and T, constitutively active; and C, catalytically inactive.
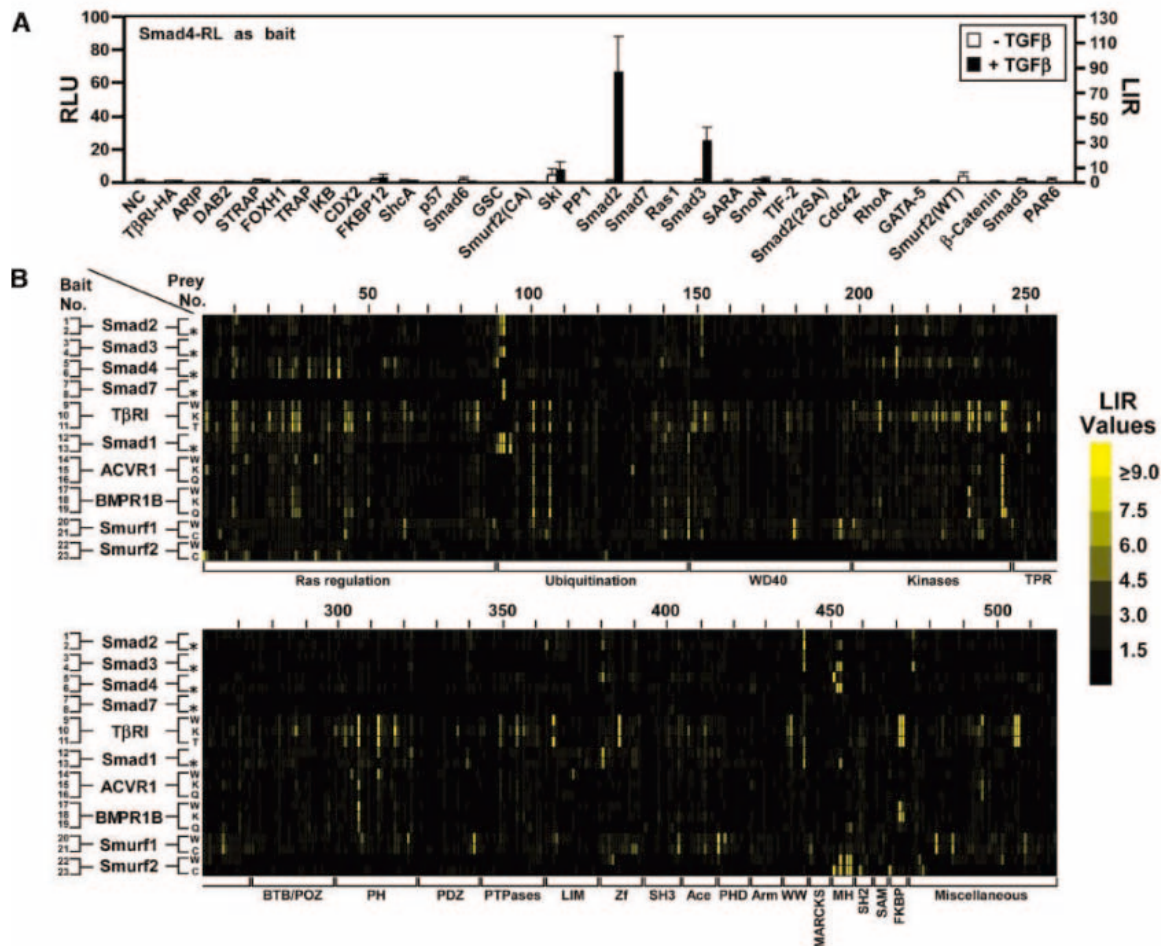
TGFβ superfamily of extracellular morphogens, which regulates a plethora of biological processes in metazoans (6, 7). TGFβs signal through ligand-induced heteromeric receptor complexes of type II and type I transmembrane serine-threonine kinases, which are activated when the type II receptor kinase transphosphorylates the type I receptor (fig. S1A). This stimulates binding and phospho-

[1]Program in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada, M5G 1X5. [2]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, M5G 2M9. [3]Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario, Canada, M5S 1A8. [4]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada, M5S 3H5. [5]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada, M5G 1L6. [6]Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan. [7]Division of Cancer Informatics, Ontario Cancer Institute, Princess Margaret Hospital, Toronto, Ontario, Canada, M5G 2M9.

*To whom correspondence should be addressed. E-mail: wrana@mshri.on.ca

rylation of receptor-regulated Smads (R-Smads) by the type I receptor. Phosphorylated R-Smads then dissociate from the receptor and form a complex with Smad4, and this complex then accumulates in the nucleus where it regulates transcription by interacting with DNA binding proteins (7, 8).

The Smad pathway provides an example of how posttranslational modifications (PTMs) regulate the dynamics of PPI networks to control signal transduction (9). Therefore, we used this pathway to determine whether LUMIER could map PTM-dependent interactions. For this, we fused RL to Smad4 (Smad4-RL) and coexpressed it with Flag-Smad2 or the TGFβ receptor phosphorylation site mutant, Flag-Smad2(2SA), which does not bind Smad4. In the absence of signaling, little if any Smad4-RL was found associated with Smad2 (Fig. 1B), whereas TGFβ signaling induced a strong association with wild-type (WT) Smad2 that was revealed by high levels of RL activity in the immune complexes (10). In contrast, no Smad4-RL was detected in Smad2(2SA) immunoprecipitates. Similar results were

obtained when we switched the tags for Smad2 and Smad4 (11). LUMIER also detected signal-specific interactions in the TGFβ-related bone morphogenetic protein (BMP) pathway (fig. S1B), constitutive interactions such as that between Smad4 and SnoN (fig. S1C), and transient substrate-enzyme interactions between Smad2 and active TGFβ type I receptor (TβRI) (Fig. 1C), which are difficult to detect by traditional means. Next, we optimized LUMIER in a 96-well format and performed a pilot-screen of Smad4-RL against 30 different Flag-tagged cDNAs in the presence or absence of TGFβ signaling (Fig. 2A). This revealed low background and strong, signal-dependent interactions between Smad4-RL and either Flag-Smad2 or Flag-Smad3, as well as the signal-independent interaction of Smad4-RL with Flag-Ski (12).

To map a TGFβ PPI network, we RL-tagged core members of the pathway (table S1) and 3×Flag-tagged (3Flag) 518 cDNAs from the FANTOM1 library (13) (fig. S2). These cDNAs encode proteins that contain at least one of the domains summarized in Fig. 2B (10). Each tagged protein was transiently expressed in mammalian cells, and their expression and subcellular localization were characterized (fig. S3) (10, 14). To analyze the interaction of each RL-tagged TGFβ pathway component in different signaling contexts (table S1) with every Flag-tagged protein required about 12,000 experiments. Therefore, we used a robotics platform and performed automated LUMIER (10). To visualize the entire dataset, we generated a diagram in which each screen with RL-tagged protein is represented on the vertical axis and Flag-tagged proteins are on the horizontal axis (Fig. 2B). The results of each PPI test are represented by a bar at the intersection, with the intensity of yellow color reflecting that of the interaction, calculated as fold change over the negative control (10). We refer to this as the LUMIER intensity ratio (LIR) (see table S2 for all LIR values).

At low LIRs, background noise predominates, resulting in high false-positive rates and a noisy network, whereas at higher LIRs, interactions will be of higher confidence, but interactions that may be transient, occur in specific compartments, or are of lower affinity will be lost. Therefore, to build the interaction network graph, we examined false-negative rates and false-positive rates (10) and chose a conservative LIR cutoff of 3. This yielded false-negative rates of 36% (table S3) and false-positive rates of ~20%. The TGFβ interaction network at this cutoff revealed 947 interactions from 11,914 tests (table S2) that formed an interconnected network with nodes displaying degrees that ranged from 1 to 134 (table S4). The network

Fig. 3. The TGFβ interactome by LUMIER. (**A**) Network graph of the TGFβ interactome. Proteins are nodes, color-coded according to their Gene Ontology annotation (inset), and interactions with a LIR of 3 or greater are shown as blue edges. For clarity, Smads under different signaling conditions and various WT, activated, and catalytically inactive versions of Smurfs and receptors have been condensed into single nodes. (**B**) Schematic of the bait map for the SOMs showing the positioning of the baits. The numbers correspond to the bait screen as defined in Fig. 2B. Note that the distance between baits reflects the similarity in their interaction profile with the 518 preys. Thus, baits that behave similarly are close to each other and when very similar can occupy the same position. (**C**) Combined unsupervised and supervised BTSVQ clustering of the LUMIER data identifies a group of proteins with similar SOMs. IRF3 (lower right), which is part of a different cluster, is shown for comparison.

displayed features of a scale-free network with possible hierarchical modularity (10) (fig. S4), which have also been noted in other PPI networks (15). Network dynamics are important in understanding signaling networks (16), which must be remodeled in time and space to convey information. To deter-

mine how signaling regulates PPIs in our network, we focused on the dynamics of the Smad2 and Smad4 interactome in the absence and presence of TGFβ signaling (fig. S5). To capture the dynamics, we also generated a movie of the network changes (movie S1). Analysis of these graphs and the movie

revealed considerable partner switching; numerous interactions were lost upon signaling, whereas others, such as the assembly of Smad complexes, were stimulated.

LUMIER can thus detect in mammalian cells pathway-specific, PTM-dependent PPIs, constitutive protein interactions, and interactions involving transmembrane receptors. The last mentioned is noteworthy, as transmembrane receptors are critical focal points in signaling networks, are important drug targets, and have been difficult to study with other high-throughput approaches (17). However, we have not yet been able to measure the concentration of the Flag-tagged preys in high-throughput LUMIER; therefore, we cannot measure absolute PPI affinities. Furthermore, overexpression can be useful for detecting interactions that occur between proteins of low abundance, as well as weak or transient interactions, but this can also make the assay prone to false positives. These advantages and disadvantages should be taken into account when evaluating LUMIER networks.

To identify novel TGFβ signaling networks of biological importance, we explored the TGFβ LUMIER dataset using the binary tree–structured vector quantization (BTSVQ) algorithm (10, 18). BTSVQ performs unsupervised clustering and supports intuitive visualization of the data, both as a binary tree dendrogram and as Kohonen's self-organizing maps (SOMs), providing a powerful method for visualizing complex datasets that effectively exploits the semiquantitative information inherent in LUMIER. To identify novel targets of the TGFβ pathway, we focused on clustering the Flag-tagged proteins (preys). For this, the algorithm first analyzed the interaction of each RL-tagged bait with the library of Flag-tagged preys and organized the placement of baits according to their interaction profiles with all of the preys (see Fig. 3B for map of bait positions). Therefore, baits that display different interaction profiles are far apart, whereas baits displaying similar profiles are close (Fig. 3B). When bait profiles are very similar they can occupy the same map unit position. Once the positions of the baits are fixed, a SOM is generated for each prey, in which the interaction of that prey with each of the baits is projected onto the bait map, with the color representing the LIR value. Unlabeled map units are interpolated to make the color homogeneous (Fig. 3C).

BTSVQ analysis thus identified clusters of prey proteins that displayed similar patterns of interactions with the TGFβ pathway (that is, similar SOM patterns). One of these clusters included PAK1, a member of the PAK family, which is involved in regulating cytoskeletal dynamics, cell motility, survival, proliferation, and gene expression through a variety of effectors (19) and has been implicated in TGFβ signaling (20). However,
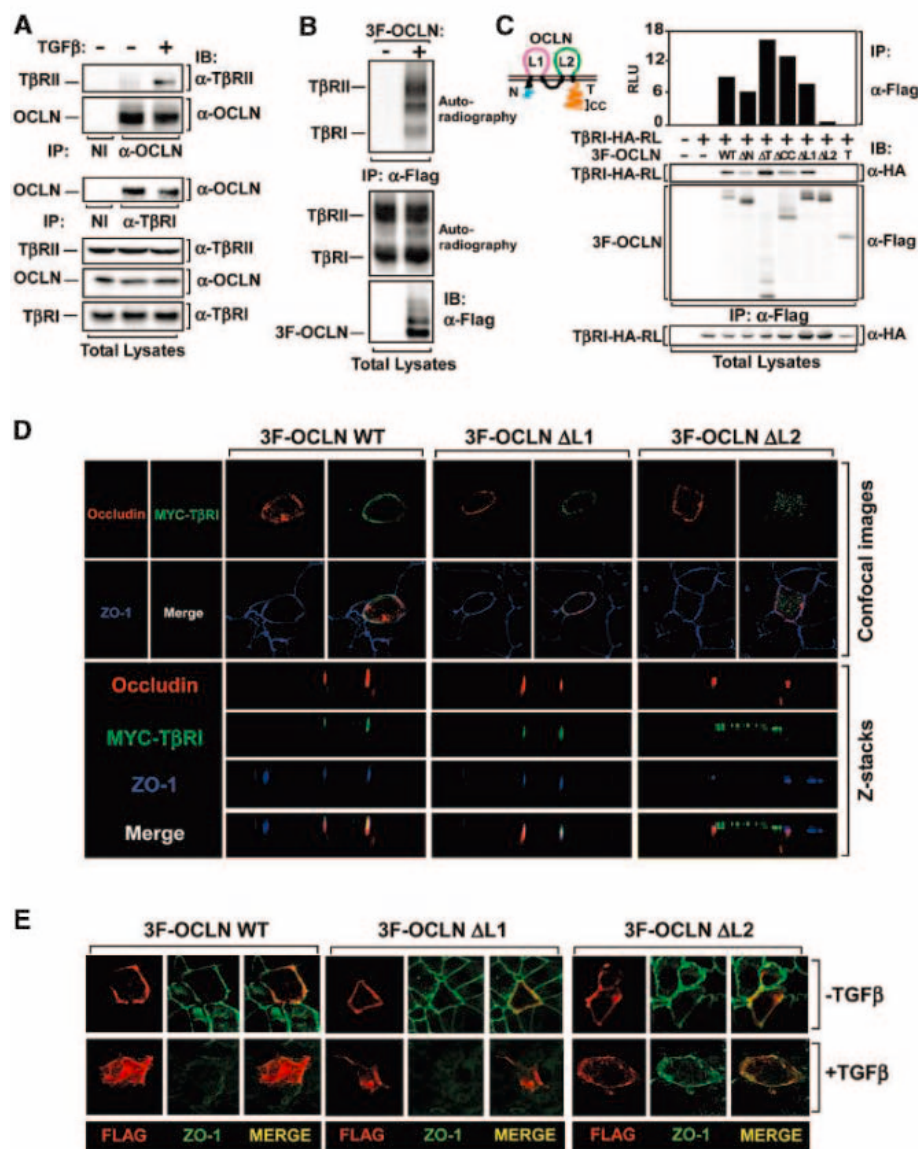
Fig. 4. OCLN regulates TGFβ-dependent dissolution of tight junctions. Interaction between endogenous OCLN and type I and type II receptors in untreated (–) or TGFβ-stimulated (+) NMuMG cells (A) or with affinity-labeled TGFβ receptors expressed on HEK-293T cells (B) was analyzed as described (10). (C) Mapping the TβRI interaction domain on OCLN using LUMIER. WT TβRI-HA-RL was expressed alone or together with WT or mutant OCLN harboring the indicated domain deletions (see schematic), and interactions were measured by LUMIER (top panel) or by immunoblotting with antibody against HA. Expression of proteins was confirmed by immunoblotting total cell lysates, as indicated. (D) Localization of OCLN, OCLN(ΔL1), OCLN(ΔL2), and cell surface TβRI in polarized NMuMG epithelial cells. WT 3Flag-tagged OCLN or the indicated mutants (red) were expressed together with Myc-tagged TβRI (green), and protein subcellular localization in tight junctions (ZO-1, blue) was visualized as described (10). The overlay shows colocalization of all three proteins as white (MERGE). (E) OCLN(ΔL2) inhibits TGFβ-induced EMT. NMuMG cells transiently expressing the indicated WT or OCLN mutants were untreated (–) or treated (+) with TGFβ for 24 hours. OCLN-expressing cells were identified by staining with antibody against Flag, and tight junctions were visualized by staining with antibodies against ZO-1. Confocal optical slices of the tight junction region are shown.

physical association with TGFβ pathway components has not been reported. Therefore, we anchored our unsupervised clustering on PAK1 and sorted prey SOMs to identify similar profiles. The resulting cluster (Fig. 3C) contained known PAK1-binding proteins, such as the PAK-interacting Cdc42 exchange factor, ARHGEF6 (α-PIX), and oxidative stress–response kinase-1 (OSR1) (19, 21), as well as Occludin (OCLN), a tight junction accessory protein that is associated with the cell polarity network (22) (fig. S7A). Therefore, we validated the interaction of a number of these proteins with TGFβ receptors (fig. S6) (10), which included validating interactions between endogenous TβRI and both PAK1 and OCLN. Further, although physical links reported between PAK1 and Polarity-OCLN networks are few (fig. S7A), analysis of LUMIER data revealed substantial connections between the TGFβ pathway and both networks (fig. S7B).

TGFβ induces dissolution of tight junctions and acquisition of a mesenchymal phenotype in breast epithelial culture models (23, 24). Therefore, we examined the OCLN-TGFβ receptor complex in normal mammary gland epithelial cells (NMuMG). Interaction of endogenous OCLN with endogenous TβRI was not modulated by TGFβ (Fig. 4A), whereas its association with the TGFβ type II receptor increased in a ligand-dependent manner (Fig. 4A) and OCLN interacted with cell surface affinity–labeled TGFβ receptor complexes (Fig. 4B). Using LUMIER, we mapped the TβRI-interacting region of OCLN to extracellular loop 2 (L2) (Fig. 4C). TβRI is localized to tight junctions in polarized NMuMG cells (25). To determine whether OCLN might contribute to regulating TβRI localization, we used OCLN(ΔL2) as a dominant negative. Confocal microscopy in polarized NMuMG cells revealed that the WT, as well as the extracellular loop 1 (ΔL1) and ΔL2 mutants of OCLN, localized with ZO-1 on the apical aspect of the cell in tight junctions (Fig. 4D). Localization of Myc-tagged TβRI in tight junctions was unaffected by WT OCLN or OCLN(ΔL1), both of which interacted with TβRI. In contrast, OCLN(ΔL2) caused mislocalization of TβRI across the surface of the cell (Fig. 4D). Moreover, when we examined the epithelial-to-mesenchymal transition (EMT) in these cells, 40% of OCLN(ΔL2)-expressing cells exhibited retained tight junctions after TGFβ treatment, compared with only 10% of control cells (Fig. 4E) (11). In contrast, none of the OCLN mutants affected TGFβ-dependent induction of a Smad-responsive reporter gene (fig. S8). Thus, OCLN regulates TβRI localization to tight junctions, and this is important for efficient TGFβ-dependent dissolution of tight junctions during EMT. This suggests that targets of the receptor complex localized to tight

junctions are involved in EMT, and in a separate study, we show that Par6, a key regulatory component of tight junctions, is an important downstream effector of this pathway (25). Our results provide a systematic analysis in mammalian cells of protein interactions involved in cell signaling and highlight how these approaches can uncover new connectivities in mammalian signaling pathways.

References and Notes

1. P. Uetz et al., Nature 403, 623 (2000).
2. A. C. Gavin et al., Nature 415, 141 (2002).
3. Y. Ho et al., Nature 415, 180 (2002).
4. L. Giot et al., Science 302, 1727 (2003).
5. S. Li et al., Science 303, 540 (2004).
6. A. B. Roberts, L. M. Wakefield, Proc. Natl. Acad. Sci. U.S.A. 100, 8621 (2003).
7. R. Derynck, Y. E. Zhang, Nature 425, 577 (2003).
8. L. Attisano, J. L. Wrana, Science 296, 1646 (2002).
9. T. Pawson, Cell 116, 191 (2004).
10. Materials and methods are available as supporting material on Science Online.
11. M. Barrios-Rodiles, B. Ozdamar, unpublished data.
12. K. Luo et al., Genes Dev. 13, 2196 (1999).
13. J. Kawai et al., Nature 409, 685 (2001).
14. All subcellular localization data and LUMIER results can be found at http://ophid.utoronto.ca/LUMIER/.
15. S. H. Yook, Z. N. Oltvai, A. L. Barabasi, Proteomics 4, 928 (2004).
16. J.-D. J. Han et al., Nature 430, 88 (2004).
17. I. Stagljar, S. Fields, Trends Biochem. Sci. 27, 559 (2002).
18. M. Sultan et al., Bioinformatics 18 (suppl. 1), S111 (2002).
19. G. M. Bokoch, Annu. Rev. Biochem. 72, 743 (2003).
20. M. C. Wilkes, S. J. Murphy, N. Garamszegi, E. B. Leof, Mol. Cell. Biol. 23, 8878 (2003).
21. W. Chen, M. Yazicioglu, M. H. Cobb, J. Biol. Chem. 279, 11129 (2004).
22. L. Gonzalez-Mariscal, A. Betanzos, P. Nava, B. E. Jaramillo, Prog. Biophys. Mol. Biol. 81, 1 (2003).
23. S. E. Seton-Rogers et al., Proc. Natl. Acad. Sci. U.S.A. 101, 1257 (2004).
24. M. Oft et al., Genes Dev. 10, 2462 (1996).
25. B. Ozdamar et al., in press.

A Transmembrane Intracellular Estrogen Receptor Mediates Rapid Cell Signaling

Chetana M. Revankar,[1,2] Daniel F. Cimino,[1,2] Larry A. Sklar,[2,3] Jeffrey B. Arterburn,[4] Eric R. Prossnitz[1,2]*

The steroid hormone estrogen regulates many functionally unrelated processes in numerous tissues. Although it is traditionally thought to control transcriptional activation through the classical nuclear estrogen receptors, it also initiates many rapid nongenomic signaling events. We found that of all G protein–coupled receptors characterized to date, GPR30 is uniquely localized to the endoplasmic reticulum, where it specifically binds estrogen and fluorescent estrogen derivatives. Activating GPR30 by estrogen resulted in intracellular calcium mobilization and synthesis of phosphatidylinositol 3,4,5-trisphosphate in the nucleus. Thus, GPR30 represents an intracellular transmembrane estrogen receptor that may contribute to normal estrogen physiology as well as pathophysiology.

Estrogen (17β-estradiol, E2) represents one of a family of steroid hormones that act through soluble intracellular receptors. Once activated, these receptors translocate to the nucleus, where they function as ligand-dependent transcription factors (1, 2). This

mode of action of two such estrogen-binding receptors, ERα and ERβ, is reasonably well understood (3, 4). However, the existence of functional ERs associated with the plasma membrane has been debated (5). It has been suggested that such membrane receptors me-

diate the rapid nongenomic signaling events widely observed following stimulation of cells and tissues with estrogen, including the generation of the second messengers Ca^{2+} and nitric oxide as well as the activation of receptor tyrosine kinase and protein-lipid kinase pathways (1, 6–9). The cellular consequences can include adhesion, migration, survival, proliferation, and cancer. Novel receptors and novel forms of ER have been postulated to mediate many of these signal transduction events (8).

G protein–coupled receptors (GPCRs) represent the largest class of signaling molecules

in the human genome (10). They are heptahelical transmembrane proteins (11) expressed on the cell surface, where the binding of agonists (12) initiates activation of heterotrimeric guanine nucleotide-binding proteins (G proteins) and downstream regulatory proteins (13) and effectors (14). It has been suggested that rapid nongenomic estrogen-mediated signaling may involve the GPCR, GPR30 (15) based on correlations of receptor expression with estrogen-mediated Erk activation (16), and cellular estrogen binding (17). Despite these assertions that a 7-transmembrane GPCR may be involved in estrogen-dependent signaling, no mechanisms have been demonstrated to explain the function of GPR30 in these processes.

To define the role of GPR30 in estrogen-mediated cell activation, we expressed GPR30 as a fusion protein with green fluorescent protein (GFP) in COS7 (monkey kidney fibroblast) cells (18). To examine the cellular localization

of the receptor, GPR30-GFP was expressed with the β_2-adrenergic receptor (β_2-AR), a GPCR (Fig. 1A). Confocal fluorescence microscopy revealed that the β_2-AR was localized to the plasma membrane, whereas GPR30-GFP was observed in an intracellular tubulo-reticular network, including the nuclear envelope in some cases. To identify the specific subcellular localization of GPR30-GFP, we performed localization studies with markers for mitochondria, the endoplasmic reticulum, the Golgi apparatus, and the actin cytoskeleton. Staining of GPR30-GFP–expressing cells with either an antibody to KDEL (an endoplasmic reticulum retrieval sequence) or er-YFP (a subcellular localization marker targeting YFP to the endoplasmic reticulum using calreticulin and KDEL sequences) yielded nearly complete colocalization with GPR30-GFP. Coexpression of TGN38-GFP (a Golgi-resident protein) with GPR30-monomeric red fluorescent protein1 (mRFP1) also demonstrated co-

[1]Department of Cell Biology and Physiology, [2]Cancer Research and Treatment Center, and [3]Department of Pathology, University of New Mexico Health Sciences Center, Albuquerque, NM 87131, USA. [4]Department of Chemistry and Biochemistry, New Mexico State University, Las Cruces, NM 88003, USA.

*To whom correspondence should be addressed. E-mail: eprossnitz@salud.unm.edu

Fig. 1. Localization of GPR30 in the endoplasmic reticulum. (A) Confocal images of GPR30-GFP (or mRFP1) coexpressed with the β_2-AR-mRFP1, an endoplasmic reticulum–targeted yellow fluorescent protein (er-YFP), or the Golgi marker TGN38-GFP in COS7 cells. The endoplasmic reticulum was also stained with an antibody to KDEL. Mitochondria and the actin cytoskeleton were visualized with Mito-Tracker Red CMXRos (Molecular Probes, Eugene, OR) and rhodamine-phalloidin, respectively. (B and C) Confocal images of full-length GPR30-GFP (B) or N-terminally FLAG-tagged, C-terminally truncated GPR30 (amino acids 1 to 349) expressed in COS7 cells (C). Antibodies to N-terminal GPR30, C-terminal GPR30, and M2 FLAG were used to stain both of the transfected cells. (D) N-terminally FLAG-tagged, C-terminally truncated GPR30 was expressed in Hec50 cells, which express endogenous GPR30, and stained with antibodies to C-terminal GPR30 and M2 FLAG. (E) Confocal images of cell lines stained with antibody to C-terminal GPR30 to determine endogenous GPR30 expression levels. All images were acquired with identical instrument settings at one sitting. (F) HTR8 (green) and JEG (red) cells were stained with antibody to C-terminal GPR30 (solid line) or secondary antibody alone (dotted line) and analyzed by flow cytometry for fluorescence intensity. HTR8 cells yield a 1.5-fold increase in fluorescence intensity above background, whereas JEG cells demonstrate a 15-fold increase. Data are representative of three independent experiments. In images where the location of the nucleus is not clear, the merged image includes the TO-PRO-3 staining of the nucleus (blue).
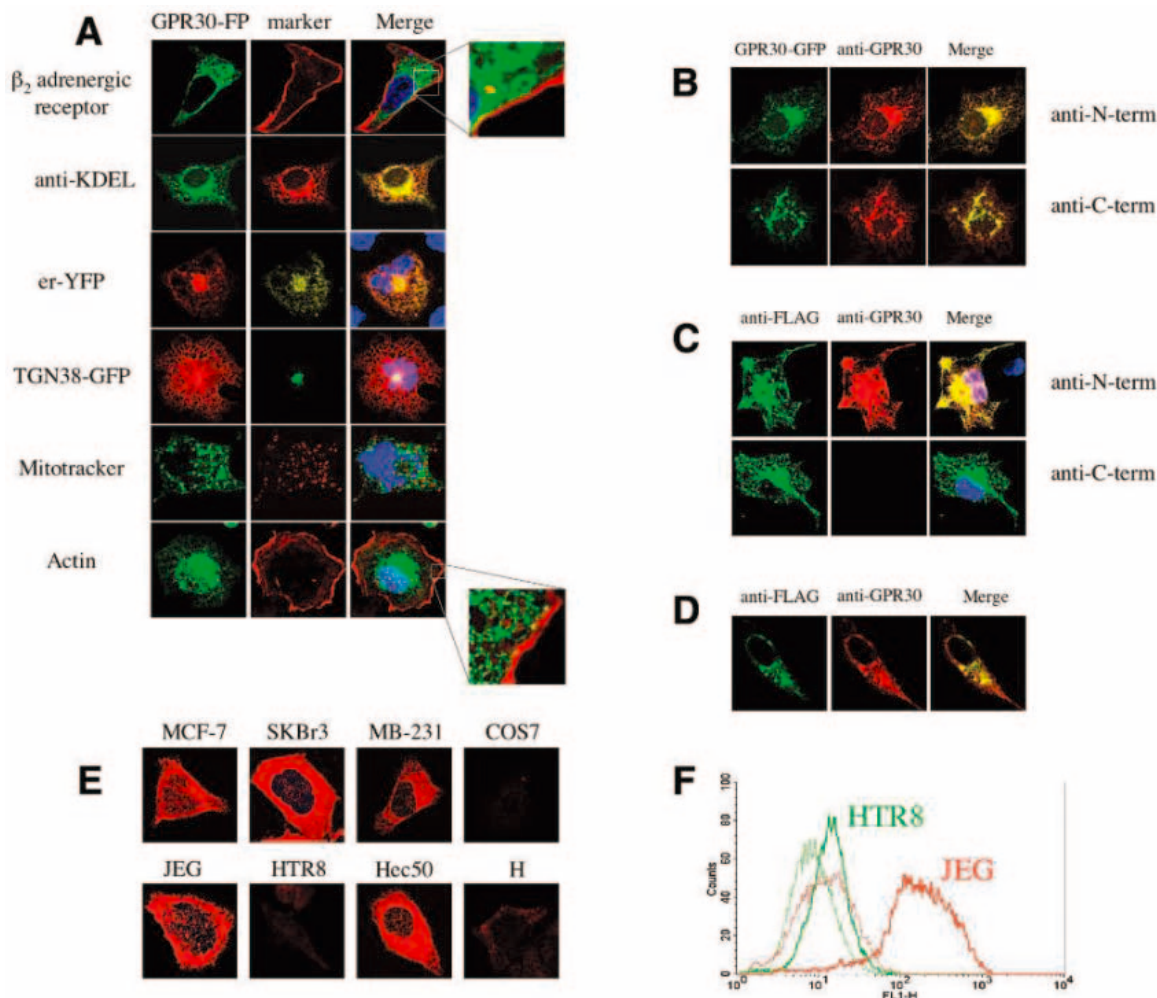
localization. No colocalization was observed between GPR30-GFP and mitochondria or the actin cytoskeleton.

Because heterologous expression of GPCRs can result in retention of the receptor in the endoplasmic reticulum and Golgi, we examine the expression of endogenous GPR30. Antibodies were generated against the predicted N- and C-terminal peptides of GPR30, both of which recognized GPR30-GFP expressed in COS7 cells (Fig. 1B). The C-terminal antibody did not detect an amino-terminally epitope-tagged truncated form of GPR30 that lacked the C-terminal 24 amino acids (Fig. 1C), although it was still expressed in the endoplasmic reticulum. Both the C-terminally truncated form of GPR30 and endogenous GPR30 were expressed with the same localization pattern (Fig. 1D). Furthermore, this C-terminal antibody identified endogenous GPR30 in a number of cancer cell lines [breast cancers MCF7, SKBr3, and MDA-MB231; juxtaglomerular epithelioid granular (JEG) choriocarcinoma; and Hec50 uterine carcinoma] in a localization pattern identical to that observed for GPR30-

GFP in transfected COS7 cells (Fig. 1E). MDA-MB-231 breast cancer cells exhibited somewhat lower expression of GPR30, whereas HTR8 normal extravillous trophoblasts, Ishikawa H cells (differentiated hormone-responsive uterine cells), and COS7 cells exhibited little to no expression. Whereas the aggressive cancer cell lines (JEG and Hec50) expressed high levels of GPR30, their associated normal cell lines (HTR8 and H, respectively) showed little to no expression of GPR30. JEG cells expressed a level of GPR30 at least 10 times as high as that of HTR8 cells (Fig. 1F), quantitatively confirming the results observed by confocal microscopy (Fig. 1E). Taken together, these results demonstrate that GPR30 is a 7-transmembrane receptor localized predominantly in the endoplasmic reticulum.

Intracellular localization has important implications for the function of a GPCR. Any natural ligand for the receptor must be membrane permeable and, furthermore, signal transduction molecules (e.g., G proteins and kinases) accessible to the receptor on internal membranes may be very different from those

available to a receptor on the plasma membrane. To date, a cognate stoichiometric ligand for GPR30 has not been firmly established. Because GPR30-positive cells were responsive to estrogens, we determined whether GPR30 could directly bind to estrogen. We developed a family of fluorescent estrogens based on 17α-[4-aminomethyl-phenylethynyl]-estra-1, 3,5(10)-triene-3, 17β-diol, which contains a single reactive amine at the distal end of a linker that is attached to the 17α position of 17β-estradiol (19). This estrogen derivative was conjugated to the N-hydroxysuccinimidyl ester forms of various Alexa dyes (Fig. 2A). The resulting products, E2-Alexa 546 and E2-Alexa 633 were purified by reverse-phase high-performance liquid chromatography. To test the binding ability of these fluorescent estrogens, we expressed ERα-GFP in COS7 cells and stained the transfected cells with E2-Alexa 633. Unpermeabilized cells did not stain with E2-Alexa 633 (or 546) because of the impermeability of the charged Alexa dyes (Fig. 2B). However, transfected cells permeabilized with saponin demonstrated binding that coincided

Fig. 2. Binding of a fluorescent estrogen analog to GPR30. (**A**) Structure of Alexa 546-modified estrogen derivative. (**B** and **C**) Confocal images of (B) ERα-GFP–expressing or (C) GPR30-GFP–expressing cells stained with E2 Alexa 546 or 633 under nonpermeabilized or saponin-permeabilized conditions in the absence or presence of excess competing 17β-estradiol (1 μM). In (B), the cell membrane in the merged image is shown by virtue of coexpression of β_2-AR-mRFP1. (**D**) Confocal images of the same eight cell lines shown in Fig. 1E but now stained with E2 Alexa 546 following permeabilization with saponin. All images were acquired with identical instrument settings at one sitting. (**E**) ERα, ERβ, GPR30 and β_2-AR GFP chimerae were expressed in COS7 cells and stained with E2 Alexa 633. Various receptor expression levels were defined by the use of markers in FL1 (GFP) and analyzed for the corresponding level of E2 Alexa 633 binding (specific binding was determined in the presence of 1 μM 17β-estradiol). (**F**) GPR30-GFP was expressed in COS7 cells and stained with 2 nM E2 Alexa 633 in the presence of the indicated concentration of 17β-estradiol. GPR30-GFP–positive cells were selected by a marker in FL1 (GFP) and analyzed for specific E2 Alexa 633 binding (as above). Data are representative of three independent experiments [(B) to (D)] or represent the mean ± SEM from three experiments [(E) and (F)].
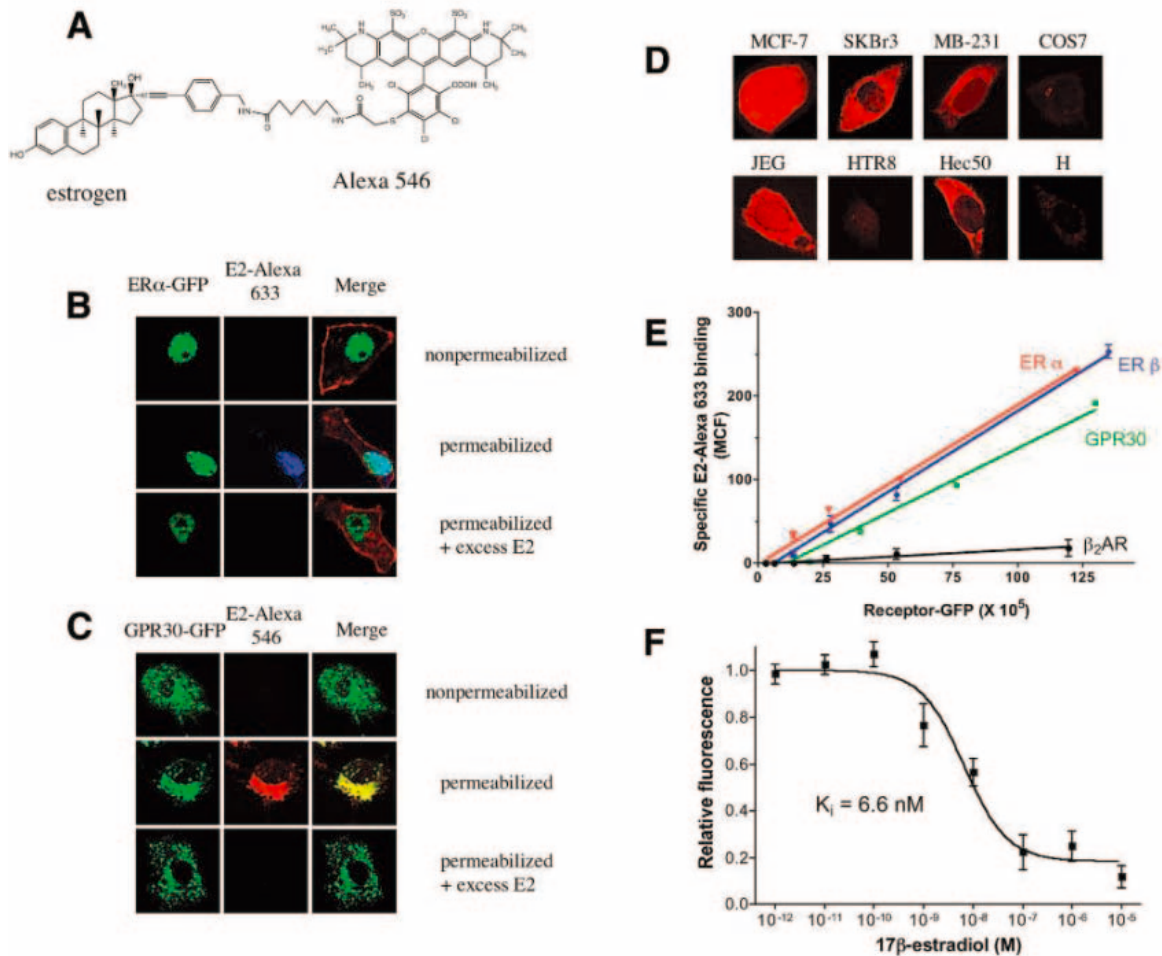
with the nuclear localization of ERα-GFP. Specificity of the binding to ERα was demonstrated by the ability of an excess of 17β-estradiol to compete for the binding of the E2-Alexa derivatives. Confocal fluorescence microscopy revealed that E2-Alexa 546 staining also colocalized nearly completely with GPR30-GFP expression in the endoplasmic reticulum and Golgi and that the binding could be competed by excess 17β-estradiol (Fig. 2C). No binding of the E2-Alexa 546 to the plasma membrane was observed, confirming the absence of GPR30 at the plasma membrane. We next stained the same panel of cell lines used in Fig. 1E with E2-Alexa 546 (Fig. 2D). In each cell type, the intensity and pattern of staining of the fluorescent estrogen was consistent with the level and pattern of GPR30 protein expression (Fig. 1E). To characterize the binding properties of the novel fluorescent estrogens further, we compared the level of E2-Alexa 633 binding to the expression level of ERα-GFP,

ERβ-GFP, and GPR30-GFP in transfected COS7 cells, using β$_2$-AR-GFP–expressing cells as a control. A direct linear correlation was observed, which supports direct binding of estrogen by GPR30 (Fig. 2E). Competition binding assays of E2-Alexa 633 binding with 17β-estradiol demonstrated a K$_i$ for 17β-estradiol of approximately 6.6 nM for GPR30 (Fig. 2F). These results demonstrate that a classic GPCR superfamily member directly binds a sex steroid hormone and that GPR30 is an estrogen-binding receptor.

To determine whether GPR30 mediates rapid nongenomic signaling, we expressed GPR30-GFP and ERα-GFP in COS7 cells and then loaded the cells with the Ca^{2+} indicator Indo1-AM. Whereas 17β-estradiol stimulation of mock-transfected cells yielded only a negligible mobilization of Ca^{2+} (Fig. 3B), stimulation of either ERα-GFP–expressing or GPR30-GFP–expressing cells resulted in a mobilization of intracellular calcium similar in magnitude to the

response evoked by stimulation of purinergic receptors with adenosine triphosphate (Fig. 3, A and B). In contrast, an inactive isomer of estrogen, 17α-estradiol, did not induce a calcium response via either receptor when used at concentrations 1000 times as high as that of 17β-estradiol. Estrogen-mediated mobilization of intracellular Ca^{2+} was observed at 17β-estradiol concentrations below 0.1 nM, with an EC$_{50}$ value of approximately 0.5 nM (Fig. 3C). These results suggest that the selectivity of GPR30 toward estrogen is similar to that of the ER. This is not unexpected, because the two hydroxyl groups of estrogen represent natural sites for ligand recognition and binding (20).

As numerous estrogen-mediated rapid signaling events are sensitive to pertussis toxin, which implies the involvement of Gi/o proteins, we determined whether ERα and GPR30 stimulated calcium mobilization in this pathway. Whereas pertussis toxin completely blocked ERα-mediated calcium mobilization (Fig. 3A),
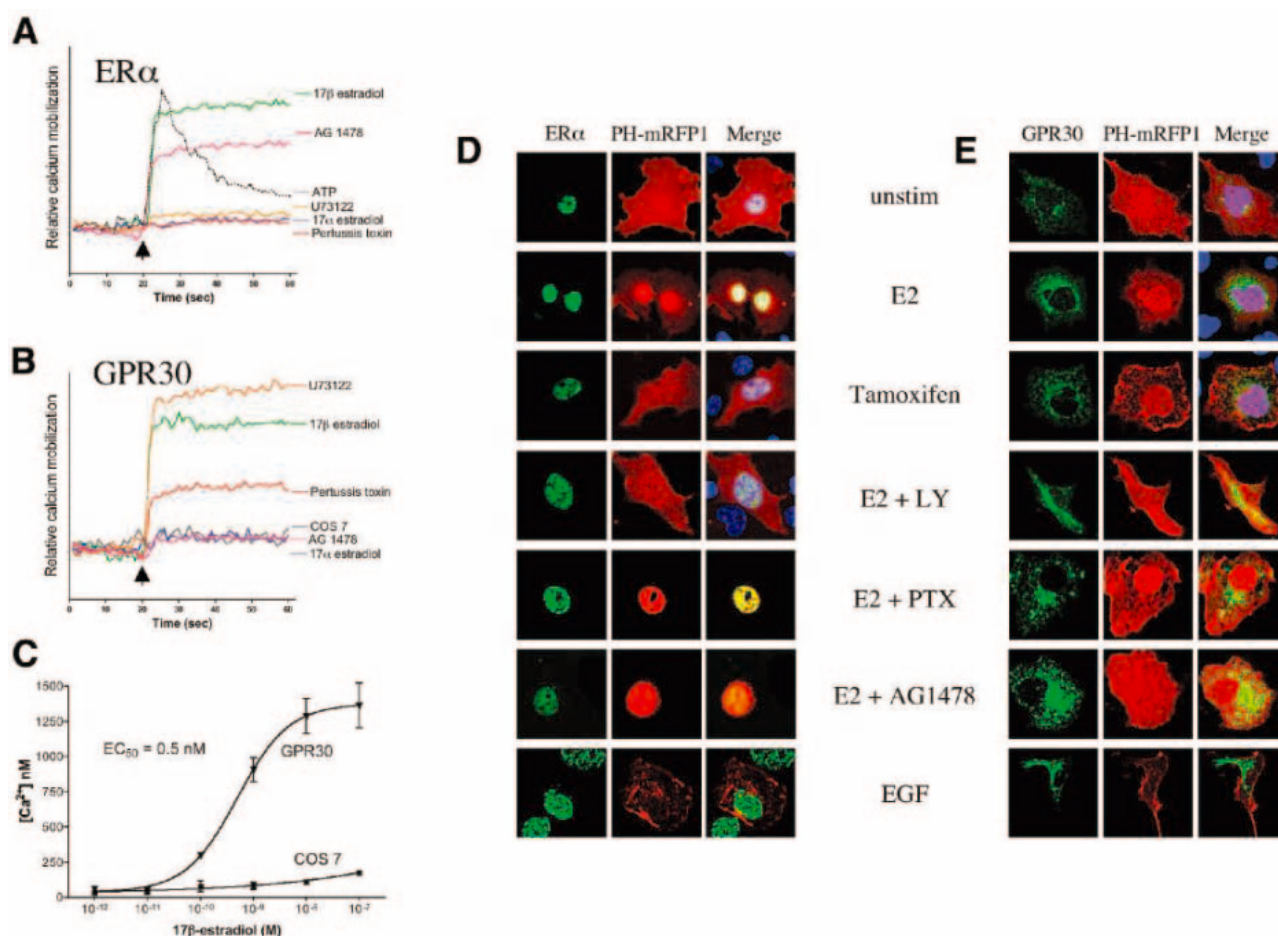
Fig. 3. Estrogen-mediated calcium mobilization and PI3K activation by GPR30 and ERα. (**A** and **B**) Intracellular calcium mobilization of Indo1-AM–loaded ERα-GFP–expressing or GPR30-GFP–expressing COS7 cells. The scales, though relative, are identical to each other in (A) and (B). In (A) and (B), 1 nM 17β-estradiol was used to stimulate the cells treated with the indicated inhibitors. In (B), COS7 refers to mock-transfected cells stimulated with 1 nM 17β-estradiol. (**C**) GPR30-GFP–expressing COS7 cells loaded with Indo1-AM were stimulated with the indicated concentration of 17β-estradiol. Mock GFP-expressing COS7 cells were used as a control (COS7). (**D** and **E**) Activation of PI3K was assessed by translocation of a PH-mRFP reporter in ERα-GFP–expressing or GPR30-GFP–expressing COS7 cells. In some merged images, the location of the nucleus is shown by TO-PRO-3 staining (blue). In all panels, the ligand concentrations used were 17β-estradiol (E2, 1 nM), 17α-estradiol (1 μM), 4-hydrotamoxifen (1 μM), EGF (100 ng/mL), and 1 nM 17β-estradiol for cells pretreated with inhibitors. Data are representative of three independent experiments [(A), (B), (D), and (E)] or the mean ± SEM from three experiments (C).

inhibition of GPR30-mediated calcium mobilization was only partial (~50%) (Fig. 3B). Furthermore, because activation of mitogen-activated protein kinase signaling by GPR30 was suggested to occur through epidermal growth factor receptor (EGFR) transactivation (16) and GPCR-mediated calcium mobilization is mediated by phospholipase C (PLC)–dependent inositol 1,4,5-trisphosphate production, we assessed whether either of these pathways is involved in estrogen-stimulated calcium mobilization by ERα and GPR30. GPR30-mediated calcium mobilization was completely blocked by the EGFR inhibitor AG1478 but not by the PLC inhibitor U73122, whereas the ERα response was completely blocked by the PLC inhibitor but not substantially by the EGFR inhibitor. These results suggest that ER- and GPR30-initiated calcium mobilization are mediated by divergent signaling pathways.

Estrogen stimulates numerous intracellular kinase signaling pathways (8), including the phosphoinositide 3-kinase (PI3K)–Akt pathway, which leads to proliferative signaling. Such estrogen-dependent signaling has been observed even in apparently ER-negative cell lines (21). Activation of PI3K leads to the membrane-localized accumulation of phosphatidylinositol 3,4,5-trisphosphate (PIP3), which in turn acts to recruit Akt to the membrane through its pleckstrin homology (PH) domain. Molecular fusion of the isolated PH domain of Akt to fluorescent proteins has been widely used to assess PIP3 accumulation and localization (22). To investigate estrogen-mediated activation

of PI3K, we transfected COS7 cells with a PH domain fused to monomeric red fluorescent protein1 (PH-mRFP1) and either ERα-GFP or GPR30-GFP. In unstimulated cells, the PH-mRFP1 reporter was localized primarily throughout the cytoplasm, with slightly higher concentrations found in the nucleus of some cells (Fig. 3, D and E). Upon stimulation with 17β-estradiol, ER-GFP–expressing cells displayed an almost complete translocation of PH-mRFP1 to the nucleus (Fig. 3D). A similar response was seen with GPR30-GFP–expressing cells (Fig. 3E), although the translocation on average was not as complete. Inhibition of PI3K by LY294002 prevented the translocation of the PH-mRFP1 to the nucleus by both receptors. Although tamoxifen [a selective estrogen receptor modulator (SERM)] induced calcium mobilization in both ERα-GFP–transfected and GPR30-GFP–transfected cells (23), it induced relocalization of the PH-mRFP1 only in GPR30-GFP–expressing cells. Furthermore, although PH-mRFP1 translocated to the nucleus, it also translocated to membranes of the endoplasmic reticulum containing GPR30-GFP. This result suggests that tamoxifen acts differentially on GPR30- and ER-mediated nuclear signal transduction and that tamoxifen activates GPR30 in a spatially different manner from estrogen. Accumulation of PIP3 by either receptor was not inhibited by pertussis toxin, indicating that ER-mediated calcium mobilization and PI3K activation occur by distinct mechanisms. PIP3 generation by GPR30, but not by ER, requires EGFR activation, indi-

cating that EGFR transactivation is similarly involved for both ER and GPR30 with respect to calcium mobilization and PI3K activation.

We examined whether activation of PI3K by intracellular GPR30 also occurs in a breast cancer cell line (SKBr3) expressing endogenous GPR30 but lacking ERα and ERβ. PH-mRFP1-transfected SKBr3 cells responded to 17β-estradiol and tamoxifen stimulation with nuclear accumulation of PIP3, as demonstrated by translocation of the PH-mRFP1 (Fig. 4A). Unstimulated cells displayed PH-mRFP1 in the plasma membrane and cytoplasm. SKBr3 cells exhibit constitutive activation of ErbB2 signaling (24), which presumably induces PI3K-dependent plasma membrane localization of PH-mRFP1 in unstimulated cells. This is supported by the cytosolic localization of the PH-mRFP1 reporter upon treatment of unstimulated cells with PI3K and EGFR inhibitors (23). These same inhibitors also prevented the nuclear accumulation of PH-mRFP1 in estrogen-stimulated cells, which demonstrates that the estrogen-mediated signaling in SKBr3 cells occurs through the EGFR and PI3K, as demonstrated for GPR30 in COS7 cells. To demonstrate that GPR30 alone was responsible for the PI3K signaling observed in SKBr3 cells, we employed antisense-mediated depletion of endogenous GPR30. Expression of GPR30 antisense resulted in a marked reduction in GPR30 expression, whereas expression of β2-AR antisense did not affect GPR30 expression (Fig. 4B). Expression of GPR30 antisense also reduced binding of E2-Alexa 546

Fig. 4. Estrogen-mediated activation of PI3K in SKBr3 breast cancer cells. (**A**) SKBr3 cells expressing PH-mRFP1 were stimulated with buffer alone (unstim), 1 nM 17β-estradiol (17β E2), 1 μM tamoxifen, 1 μM 17α-estradiol (17α E2), or 1 nM 17β-estradiol for cells pretreated with LY294002 (E2 + LY) or AG1478 (E2 + AG). (**B** and **C**) SKBr3 cells were cotransfected with either a GPR30 or β2-AR antisense construct [receptor open reading frame cloned in the reverse orientation in pcDNA3.1/Hygro(–)] and actin-GFP at a molar ratio of 5:1. Following 3 days growth, cells were permeabilized and stained with either antibody to C-terminal GPR30 (B) or E2-Alexa 546 (C). (**D**) GPR30 protein expression levels and ligand-binding capacity of GPR30 antisense-expressing cells (GPR30 AS) and control β2-AR antisense-expressing cells (β2AR AS), prepared as in (B) and (C), were quantitated by flow cytometry. In each sample, transfected cells were gated on FL1 by virtue of actin-GFP expression and compared with untransfected (nongreen) cells, providing direct internal controls. (**E**) SKBr3 cells were cotransfected with GPR30 or β2-AR antisense constructs as in (B) and (C), along with PH-mRFP1 and GFP (to mark transfected cells), to assess PI3K activation and PIP3 accumulation. Only cells depleted of GPR30 fail to respond to estrogen stimulation with translocation of the PH-mRFP1 to the nucleus (GPR30 antisense + E2 versus β2-AR + E2). Data are representative of three independent experiments [(A) to (C) and (E)] or the mean ± SEM from three experiments (D).
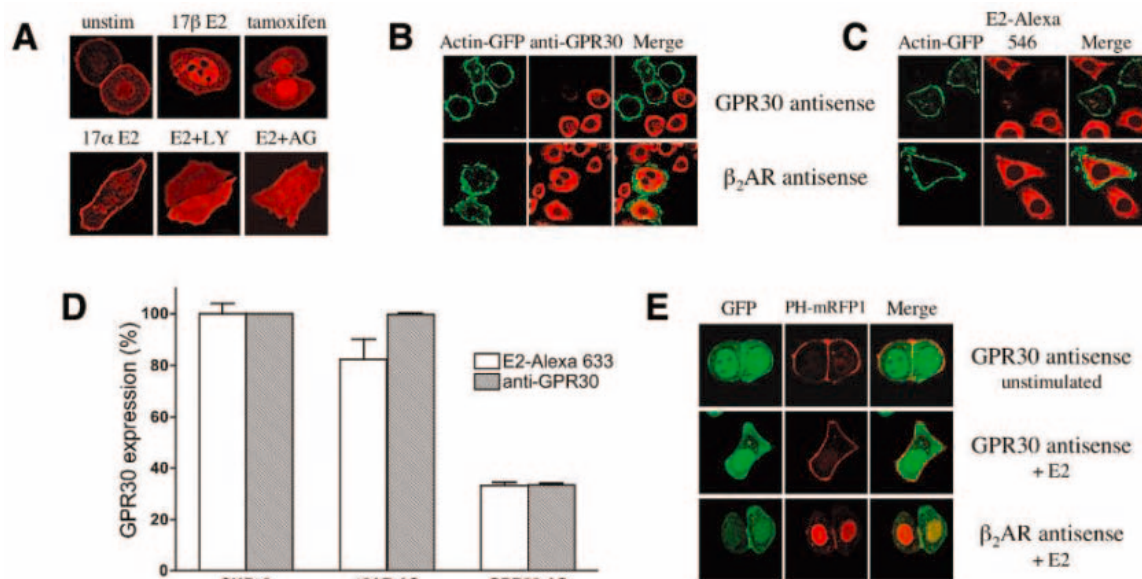
to transfected SKBr3 cells (Fig. 4, C and D). To determine whether GPR30 depletion would abrogate estrogen-mediated PI3K activation, we coexpressed antisense vectors (either GPR30 or β_2-AR), PH-mRFP1 (to assess PIP3 synthesis and localization), and GFP (to mark transfected cells). 17β-estradiol stimulation of GPR30 antisense-expressing cells did not cause PH-mRFP1 translocation (Fig. 4E), which demonstrates that the GPR30 represents the sole estrogen-binding and functionally responsive estrogen receptor in SKBr3 cells.

Mechanisms of estrogen-mediated cellular activation have become increasingly complex. Rapid nongenomic ER-mediated signaling has been proposed to occur through distinct cellular localization of the classical ER, through classical heterotrimeric G proteins, and through many of the effectors traditionally associated with growth factors and GPCRs. Estrogen is one of the small number of membrane-permeable physiological ligands, consistent with the lack of requirement for GPR30 to be expressed on the cell surface. In some cell types, GPR30 represents the sole estrogen-responsive receptor. In direct comparison with ER-mediated signaling, GPR30-mediated signaling occurs via a distinct signal transduction pathway, and the effects of estrogen through this receptor may likely be a consequence of its intracellular localization.

References and Notes

1. J. M. Hall, J. F. Couse, K. S. Korach, *J. Biol. Chem.* **276**, 36869 (2001).
2. W. Shao, M. Brown, *Breast Cancer Res.* **6**, 39 (2004).
3. S. Nilsson et al., *Physiol. Rev.* **81**, 1535 (2001).
4. D. P. McDonnell, J. D. Norris, *Science* **296**, 1642 (2002).
5. A. P. Govind, R. V. Thampan, *Mol. Cell. Biochem.* **253**, 233 (2003).
6. M. J. Kelly, E. R. Levin, *Trends Endocrinol. Metab.* **12**, 152 (2001).
7. E. R. Levin, *Steroids* **67**, 471 (2002).
8. K. J. Ho, J. K. Liao, *Arterioscler. Thromb. Vasc. Biol.* **22**, 1952 (2002).
9. M. Razandi, A. Pedram, S. T. Park, E. R. Levin, *J. Biol. Chem.* **278**, 2701 (2003).
10. J. C. Venter et al., *Science* **291**, 1304 (2001).
11. T. H. Ji, M. Grossmann, I. Ji, *J. Biol. Chem.* **273**, 17299 (1998).
12. U. Gether, B. K. Kobilka, *J. Biol. Chem.* **273**, 17979 (1998).
13. E. R. Prossnitz, *Life Sci.* **75**, 893 (2004).
14. H. E. Hamm, *J. Biol. Chem.* **273**, 669 (1998).
15. E. J. Filardo, *J. Steroid Biochem. Mol. Biol.* **80**, 231 (2002).
16. E. J. Filardo, J. A. Quinn, K. I. Bland, A. R. Frackelton Jr., *Mol. Endocrinol.* **14**, 1649 (2000).
17. P. Thomas, Y. Pang, E. J. Filardo, J. Dong, *Endocrinology* **146**, 624 (2005).
18. Materials and methods are available as supporting material on *Science* Online.
19. J. B. Arterburn, K. V. Rao, M. C. Perry, *Tetrahedron Lett.* **41**, 839 (2000).
20. A. M. Brzozowski et al., *Nature* **389**, 753 (1997).
21. E. M. Tsai, S. C. Wang, J. N. Lee, M. C. Hung, *Cancer Res.* **61**, 8390 (2001).
22. T. Balla, P. Varnai, *Science STKE* **2002**, PL3 (2002).
23. C. M. Revankar et al., data not shown.
24. K. M. Nicholson, C. H. Streuli, N. G. Anderson, *Breast Cancer Res. Treat.* **81**, 117 (2003).
25. This work was supported by NIH grant AI36357 and a University of New Mexico Cancer Research and Treatment Center Translational Research Grant to E.R.P. and by NIH grant EB00264 to L.A.S. We thank S. Kamat for excellent technical assistance and K. Leslie for helpful discussions. Flow cytometry data and confocal images in this study were generated in the Flow Cytometry and Fluorescence Microscopy Facilities, respectively, at the University of New Mexico Health Sciences Center, which received support from NCRR 1 S10 RR14668, NSF MCB9982161, NCRR P20 RR11830, NCI R24 CA88339, the University of New Mexico Health Sciences Center, and the University of New Mexico Cancer Center.

Differential Lysosomal Proteolysis in Antigen-Presenting Cells Determines Antigen Fate

Lélia Delamarre, Margit Pack, Henry Chang, Ira Mellman,* E. Sergio Trombetta*

Antigen-presenting cells (APCs) internalize antigens and present antigen-derived peptides to T cells. Although APCs have been thought to exhibit a well-developed capacity for lysosomal proteolysis, here we found that they can exhibit two distinct strategies upon antigen encounter. Whereas macrophages contained high levels of lysosomal proteases and rapidly degraded internalized proteins, dendritic cells (DCs) and B lymphocytes were protease-poor, resulting in a limited capacity for lysosomal degradation. Consistent with these findings, DCs in vivo degraded internalized antigens slowly and thus retained antigen in lymphoid organs for extended periods. Limited lysosomal proteolysis also favored antigen presentation. These results help explain why DCs are able to efficiently accumulate, process, and disseminate antigens and microbes systemically for purposes of tolerance and immunity.

Whereas antibodies typically recognize intact antigens, T lymphocytes recognize proteins as short peptides presented by the major histocompatibility complex (MHC) molecules to induce antigen-specific responses to foreign

Department of Cell Biology and Department of Immunobiology, Ludwig Institute for Cancer Research, Yale University School of Medicine, 333 Cedar Street, Post Office Box 208002, New Haven, CT 06520–8002, USA.

*To whom correspondence should be addressed. E-mail: sergio.trombetta@yale.edu (E.S.T); ira.mellman@yale.edu (I.M.)

proteins and tolerance to self proteins. Peptides generated by lysosomal proteases bind to MHC class II molecules, which are then displayed at the cell surface of a limited subset of antigen-presenting cells (APCs) including macrophages (MØs), dendritic cells (DCs), and B cells. Due in part to their marked endocytic and phagocytic capacity, MØs were long considered the prototypical APCs. It is now well established, however, that DCs are also highly endocytic and even more potent APCs (*1–3*).

Much of what is known about antigen processing is derived from studies of cultured cells. To evaluate the relative contribution of lysosomal proteases to antigen presentation in vivo, we analyzed by immunofluorescence microscopy the expression of individual proteases in situ in secondary lymphoid organs of mice. Cathepsin S (cat S), considered to be selectively expressed by APCs such as B cells and DCs, was expressed most abundantly in the MØ-rich marginal zone of the spleen (SIGN-R1-positive). Under the same staining conditions, cat S was barely detected in B cell areas or in T cell areas where the majority of DCs reside (Fig. 1A). The same pattern of expression was found for an array of other lysosomal proteases [cat L, K, and asparagine endopeptidase (AEP) (Fig. 1A) and cat B, D, E, H, and O (fig. S1)] in the spleen, whether assayed by immunofluorescence on tissue sections (Fig. 1A), on dissociated spleen cells (fig. S2), or by immunoblot (Fig. 1B). Similarly, lysosomal proteases in lymph nodes were also greatly enriched in SIGN-R1-positive MØ regions but present at low amounts in B and T cell zones where the majority of MHC class II–high cells reside (Fig. 1A). Thus, in vivo APCs differ broadly in their capacities for antigen degradation, with cells rich in MHC class II (B cells and DCs) containing substantially lower amounts of lysosomal proteases than MØs.

The marked differences in lysosomal protease content between DCs and MØs observed in vivo was recapitulated during differentiation of bone marrow precursors into the MØ lineage [by macrophage–colony stimulating factor (CSF)] or the DC lineage
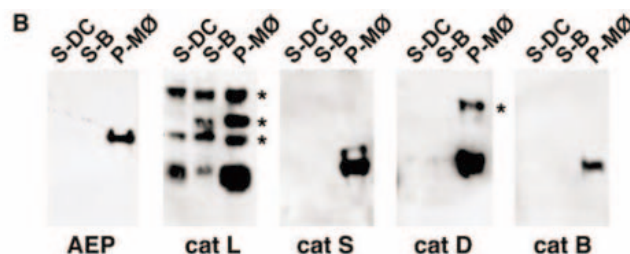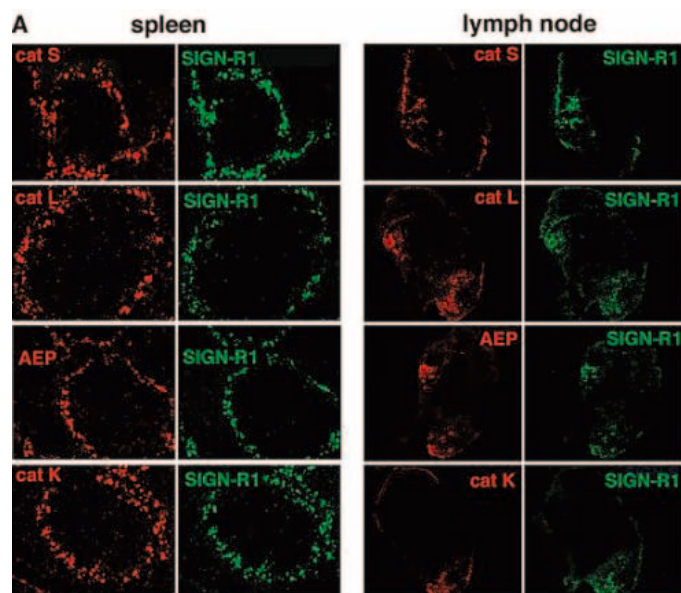
Fig. 1. DC and B cells have lower amounts of lysosomal proteases than MØs. (**A**) The indicated proteases were detected by immuno-fluorescence on frozen sections of spleen (left) and lymph node (right) from C3H/HeJ mice. AEP stands for asparagine endopeptidase or legumain. (**B**) Detection of lysosomal proteases by Western blot. DCs (S-DC) and B cells (S-B) were isolated from spleen with magnetic microbeads conjugated to a monoclonal antibody against CD11c (DCs) or CD19 (B cells). MØs (P-MØ) were isolated by adherence to plastic culture dishes from a peritoneal lavage obtained 3 days after injection of 2 ml of 3% thioglycollate. The asterisks indicate the immature, incompletely cleaved forms of the corresponding enzymes.
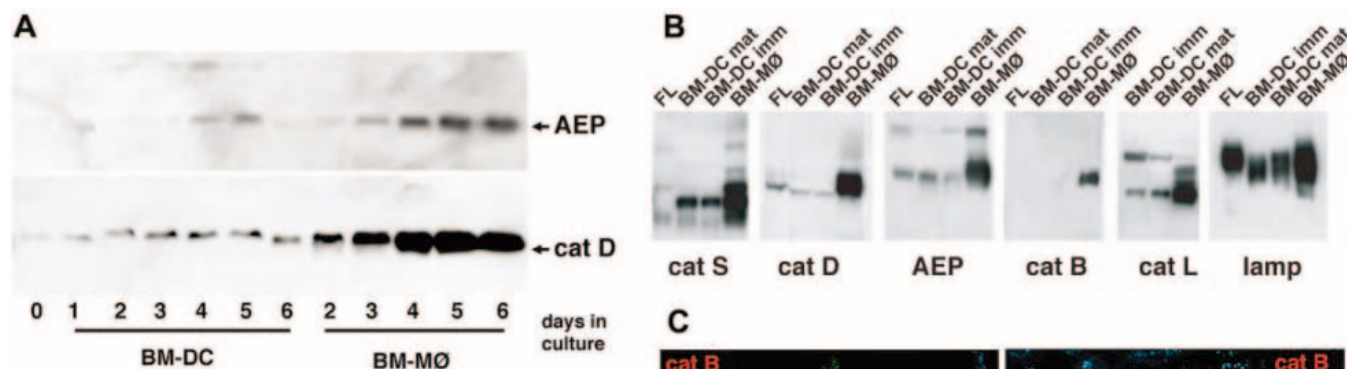
Fig. 2. Bone marrow (BM)–derived cultures of DCs and MØs recapitulate phenotype of lysosomal proteolysis observed in vivo. (**A**) Analysis of the expression of AEP and cathepsin D (cat D) during differentiation of bone marrow cultured either in GM-CSF to obtain BM-DCs (purified with use of anti-CD11c microbeads) or in M-CSF to obtain BM-MØs (purified by adherence to plastic culture dishes). Samples were taken at the indicated times of culture and analyzed by Western blot. Equal amounts of total protein (50 μg) were loaded in each lane. (**B**) Western blot analysis of the protease content from L929 fibroblasts (FL) and CD11c-positive immature or mature BM-DCs (grown in GM-CSF) and BM-MØs (grown in M-CSF). The asterisks indicate the proforms of AEP and cathepsin L (cat L). (**C**) DCs and MØs derived from bone marrow in the presence of GM-CSF were analyzed by confocal microscopy after labeling for the indicated markers. DCs are high in MHC class II, whereas MØs express abundant amounts of the indicated lysosomal proteases. MØs were identified on the basis of their low amounts of MHC class II, high amounts of Mac-3, strong adherence to glass, and high endocytic and phagocytic capacity.

(by granulocyte macrophage–CSF). Samples of DC and MØ cultures were analyzed by immunoblot for lysosomal protease content at various times during differentiation. Compared to BM-DCs, BM-MØs developed abundant enzyme amounts (Fig. 2A). Instead, DCs (immature and mature) were more similar to non-APCs such as fibroblasts, both in terms of protease content and lysosomal membrane components (lamp) (Fig. 2B). Analyzing bone marrow cultures at the single-cell level again revealed that MHC class II–high cells (DCs) were weakly stained for a variety of lysosomal proteases

(Fig. 2C). As found in situ in lymphoid organs (Fig. 1), the lysosomes in bone marrow–derived MHC class II–low MØs, on the other hand, contained high amounts of all the lysosomal proteases tested (4).

We next asked whether the disparate expression of lysosomal proteases observed in MØs, DCs, and B cells reflected differences in functional proteolytic activity. Indeed, cell-free extracts of MØs differentiated in culture (B-MØ) (Fig. 3A) or isolated from mice (by peritoneal lavage, P-MØ) (Fig. 3A) had a much higher capacity for proteolysis (at acid pH values) of casein in vitro than similar

extracts from APCs rich in MHC class II, i.e., DCs differentiated in culture [B-DC (Fig. 3A)] or DC and B cells isolated from mouse spleen [S-DC and S-B (Fig. 3A)]. Even extracts prepared from fibroblasts (FL, Fig. 3A) exhibited a greater proteolytic capacity than DCs or B cell lysosomes (Fig. 3A). The differences reflected general lysosomal proteolytic capacity and not the activity of a selected enzyme for a single substrate, because the same result was obtained for multiple other substrates including ovalbumin (OVA) (Fig. 3B), bovine serum albumin (BSA) (fig. S3), or a complex mixture of

proteins [yeast cytosol (Fig. 3C)]. A similar pattern was found in human cells isolated from human peripheral blood lymphocytes, where B cells exhibited much lower lysosomal proteolysis than monocyte-derived MØs (4). We estimated that lysosomal extracts from mouse peritoneal MØs were 20- to 60-fold more active than the corresponding extracts from mouse spleen DCs (Fig. 3D). Remarkably, this difference seemed limited to lysosomal proteases. Enzymatic assays for other lysosomal acid hydrolases (glycosidases and phosphatases) revealed comparatively minor (~twofold) differences between MØs and DCs (Fig. 3D). These smaller differences extended to the relative amounts of the lyso-somal resident membrane protein Lamp-1, which was only two to three times more abundant in MØs (Fig. 2B). Quantitative electron microscopy also revealed that DCs do not contain dramatically fewer lysosomes than MØs (4). Thus, DC lysosomes are selectively attenuated (or MØs selectively enhanced) with respect to proteolytic activity.

We then examined whether the different lysosomal proteolytic capacities of APCs affected the survival of internalized antigens in vivo by following the fate of soluble protein antigens administered to mice under different conditions. Fluorescent dextran served as a nondegradable probe to identify all endocytic APCs. Four hours after intradermal injection, APCs containing both dextran and green fluorescent protein (GFP) appeared in the draining lymph nodes, as determined by fluorescence-activated cell sorting (FACS) (table S1). The endocytic cells were predominantly MØs (CD11b+/CD11c–/class II–) and DCs (CD11c+/MHC class II–high), and both contained comparable amounts of internalized dextran and protein (4). After 20 hours, MØs remained dextran-positive, but nearly all of them were now negative for GFP, indicating that the internalized GFP had been degraded. Almost one-third of the DCs were still found to contain both dextran and GFP even after 20 hours, indicating that the half-life of GFP in DCs was substantially longer than in MØs (table S1).

These data were not restricted to GFP, because similar results were obtained for other protein antigens [horseradish peroxidase (HRP), OVA, and bovine ribonuclease A (RNase-A)] (4). We also confirmed these findings by immunofluorescence of spleens and lymph nodes in situ. When OVA was injected intravenously, 2 to 4 hours later it was predominantly associated with SIGN-R1-positive MØs in the marginal zone of the spleen (Fig. 4A). However, after 28 hours OVA was degraded and was no longer detected in MØs (Fig. 4A) but could still be observed inside CD11c-positive DCs (Fig. 4A).

Intraperitoneal injection of HRP together with dextran yielded the same results for both spleen and lymph nodes (Fig. 4B) (4). After 4 hours, most of the HRP and dextran
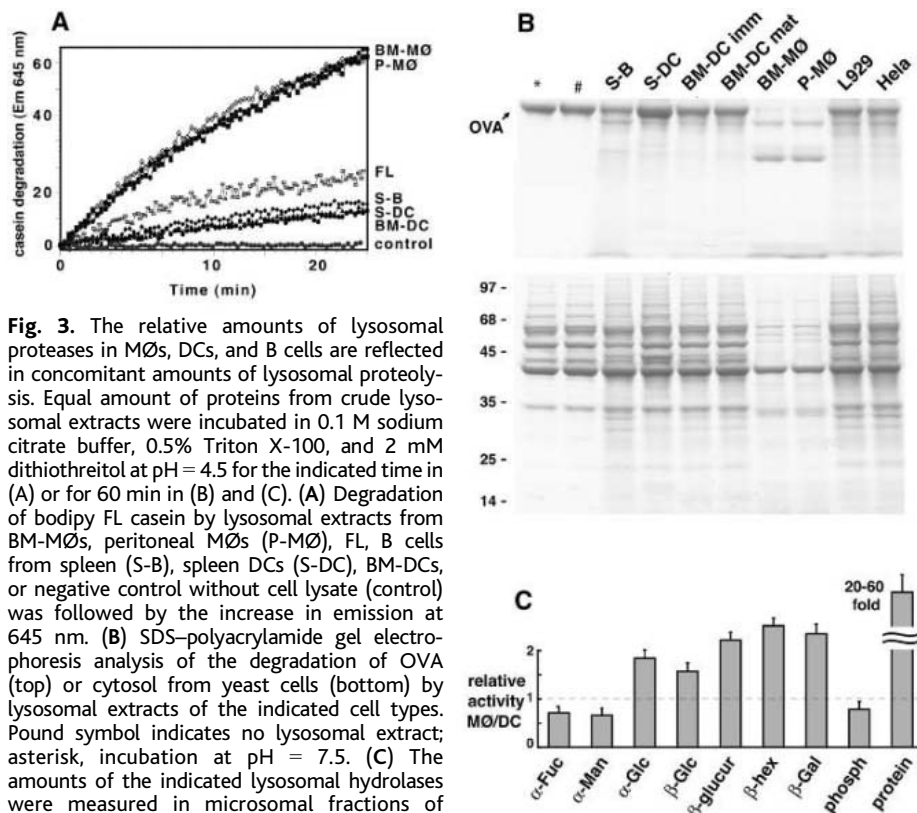
Fig. 3. The relative amounts of lysosomal proteases in MØs, DCs, and B cells are reflected in concomitant amounts of lysosomal proteolysis. Equal amount of proteins from crude lyso-somal extracts were incubated in 0.1 M sodium citrate buffer, 0.5% Triton X-100, and 2 mM dithiothreitol at pH = 4.5 for the indicated time in (A) or for 60 min in (B) and (C). (A) Degradation of bodipy FL casein by lysosomal extracts from BM-MØs, peritoneal MØs (P-MØ), FL, B cells from spleen (S-B), spleen DCs (S-DC), BM-DCs, or negative control without cell lysate (control) was followed by the increase in emission at 645 nm. (B) SDS–polyacrylamide gel electro-phoresis analysis of the degradation of OVA (top) or cytosol from yeast cells (bottom) by lysosomal extracts of the indicated cell types. Pound symbol indicates no lysosomal extract; asterisk, incubation at pH = 7.5. (C) The amounts of the indicated lysosomal hydrolases were measured in microsomal fractions of spleen DCs or peritoneal MØs and represented as their respective ratio. The results are representative of at least four independent experiments. Error bars indicate standard deviation of the mean.

was present in SIGN-R1-positive MØs (Fig. 4B). However, 18 hours after injection, the MØ retained only the undigested dextran, whereas the internalized HRP was degraded and no longer detectable (48 hours, Fig. 4B). Instead, a few CD11c-positive DCs containing both dextran and HRP were visible (Fig. 4B). Thus, independent of the antigen, the route of injection, or the ability of other APCs to internalize the administered antigen, it was the selective long-term survival of antigens in DC that appeared responsible for the long-term survival of antigens in secondary lymphoid organs.

Lastly, we asked whether the low protease activity found in DCs might not only prevent premature antigen degradation but also favor the presentation of T cell epitopes from internalized antigens. Because there are no ways to reliably modulate the amount of protease expression in DCs or MØs without also altering other important properties of these cells, we tested the hypothesis by comparing the presentation by DCs of antigens that differed only in their susceptibility to lysosomal proteases. One such antigen was RNase-A and its subtilisin-cleaved (between Ala[20] and Ser[21]) variant, RNase-S. The two forms have identical ribonuclease activities and structures (4–6). A second antigen was HRP and its destabilized form, apo-HRP, from which the bound calcium and heme

group were removed (4). The natural forms of these antigens (RNase-A and HRP) are much more resistant to proteolysis compared with their corresponding modified forms (RNase-S and apo-HRP) (4). Although all antigens (RNase-A, RNase-S, HRP, and apo-HRP) were similarly internalized by immature DCs into Lamp-positive lysosomal compartments [RNAse-A and RNase-S (Fig. 4C)], only the stable forms (RNase-A and HRP) could be easily detected in DC lysosomes after 3 hours, whereas the unstable forms [RNase-S (Fig. 4C)] and apo-HRP (4) were no longer detectable because of degradation. This difference was confirmed by biochemical measurements of protease susceptibility in vitro, where RNase-S and apo-HRP were found to be >10-fold more rapidly digested by lysosomal extracts (4). The forms that were more resistant to lysosomal digestion were more efficiently presented by APCs on MHC class II to antigen-specific polyclonal T cells isolated from mice primed with RNase-A or HRP (Fig. 4D). Thus, reduced lysosomal proteolysis actually favored the rescue of antigenic peptides and the presentation of several T cell epitopes from these two unrelated antigens.

Although perhaps counterintuitive, the finding that APCs expressing high amounts of MHC class II attenuate lysosomal proteolysis may have several biological explanations. The low amount of lysosomal proteolysis observed in

Fig. 4. Limited lysosomal proteolysis in DCs allows survival of internalized antigens in vivo. (**A**) Antigen persistence in spleen DCs after intravenous injection. OVA (8 mg) was injected intravenously into C57/B6 mice. At the indicated times after injection, OVA was detected by immunofluorescence in frozen sections of spleen. MØs and DCs were identified with use of the markers SIGN-R1 and CD11c respectively. In the top and middle images, magnification was 50× but 320× in the bottom image. (**B**) Antigen persistence in lymph node DCs after intraperitoneal injection. Similar to (A), but 3 mg of HRP were co-injected into the intraperitoneal cavity together with 0.2 mg of DNP-dextran. At the indicated times, lymph nodes were removed and stained for HRP, dextran, MØ (SIGN-R1), or DC (CD11c). (**C**) Immature BM-derived DCs were pulsed with RNase-A or RNase-S (0.5mg/ml) for 1 hour at 37°C, washed extensively, and then either fixed (top two rows) or further cultured for 3 hours (bottom two rows). MHC-II, Lamp, RNase-A, and RNase-S internalized by DCs were detected by immunofluorescence microscopy as indicated. (**D**) Limited lysosomal proteolysis enhances antigen presentation. Protease-resistant forms (RNase-A or HRP) were presented more efficiently to T cells than the forms more readily degraded (RNase-S and apo-HRP). Mouse bone marrow–derived DCs were loaded with the indicated antigens for 1 hour, washed, and then co-cultured with splenocytes from mice previously immunized with RNase-A or HRP as indicated. T cell proliferation was estimated by the incorporation of [³H]-thymidine. The results are representative of at least five independent experiments. Error bars indicate standard deviation of the mean.

DCs and B cells would render these cell types more susceptible to regulation by factors such as pH and inhibitors (7, 8). It may also help to explain why lysosomes in DCs are so efficient at recovering immunogenic peptides and assembling peptide-MHC class II complexes (9, 10). Unlike MØs, the less hostile environment found in DC lysosomes may favor the production or survival of longer peptides suitable for MHC class II association. The ability of DCs to avoid rapid degradation of internalized antigens may even contribute to their capacity to cross-present exogenous antigens on MHC class I by allowing them a greater chance to exit from endocytic organelles to the cytosol than would be expected in MØs. The limited proteolytic potential of DCs may enhance their ability to disseminate antigens throughout the immune system by minimizing the destruction of internalized antigens.

DC migration to lymph nodes can be slow (1 to 3 days), so the dissemination process would allow DCs to sequester antigens for presentation even several days later (11–14). This feature would be important for enhancing both immunogenic and tolerogenic T cell responses, because both foreign and self antigens would be treated equivalently. Indeed, the fact that commensal bacteria (15) and apoptotic cells (16, 17) can be recovered from migrating DCs may reflect the DC's limited capacity for lysosomal proteolysis. Unfortunately, pathogenic organisms, notably human immunodeficiency virus (18–21), may also make use of the comparatively safe haven provided by the DC lysosomal system, opportunistically using these efficient conveyors to reach lymphoid organs for purposes of infection.

References and Notes
1. J. Bancheeau, R. M. Steinman, *Nature* **392**, 245 (1998).
2. I. Mellman, S. J. Turley, R. M. Steinman, *Trends Cell Biol.* **8**, 231 (1998).
3. A. Lanzavecchia, F. Sallusto, *Cell* **106**, 263 (2001).
4. E. S. Trombetta, L. Delamarre, I. Mellman, unpublished observations.
5. F. M. Richards, W. W. Wickoff, Eds., *Bovine Pancreatic Ribonucleases* (Academic Press, New York, 1971), vol. IV.
6. E. E. Kim, R. Varadarajan, H. W. Wyckoff, F. M. Richards, *Biochemistry* **31**, 12304 (1992).
7. E. S. Trombetta, M. Ebersold, W. Garrett, M. Pypaert, I. Mellman, *Science* **299**, 1400 (2003).
8. S. Hartmann, R. Lucius, *Int. J. Parasitol.* **33**, 1291 (2003).
9. K. Inaba et al., *J. Exp. Med.* **188**, 2163 (1998).
10. S. J. Turley, K. Inaba, R. M. Steinman, I. Mellman, *Science* **288**, 522 (2000).
11. N. Romani et al., *J. Exp. Med.* **169**, 1169 (1989).
12. K. Inaba, J. P. Metlay, M. T. Crowley, R. M. Steinman, *J. Exp. Med.* **172**, 631 (1990).
13. M. Cella, A. Engering, V. Pinet, J. Pieters, A. Lanzavecchia, *Nature* **388**, 782 (1997).
14. P. Pierre et al., *Nature* **388**, 787 (1997).
15. A. J. Macpherson, T. Uhr, *Science* **303**, 1662 (2004).
16. T. Iyoda et al., *J. Exp. Med.* **195**, 1289 (2002).
17. C. Scheinecker, R. McHugh, E. M. Shevach, R. N. Germain, *J. Exp. Med.* **196**, 1079 (2002).
18. R. M. Steinman et al., *Curr. Top. Microbiol. Immunol.* **276**, 1 (2003).

19. A. Amara, D. R. Littman, *J. Cell Biol.* **162**, 371 (2003).
20. J. Stebbing, B. Gazzard, D. C. Douek, *N. Engl. J. Med.* **350**, 1872 (2004).
21. F. Niedergang, A. Didierlaurent, J. P. Kraehenbuhl, J. C. Sirard, *Trends Microbiol.* **12**, 79 (2004).
22. The excellent professional support provided by R. Carbone, R. Manna, and C. Marks is greatly appreciated. This work

BZR1 Is a Transcriptional Repressor with Dual Roles in Brassinosteroid Homeostasis and Growth Responses

Jun-Xian He,[1]* Joshua M. Gendron,[1,2]* Yu Sun,[1]
Srinivas S. L. Gampala,[1] Nathan Gendron,[1]
Catherine Qing Sun,[1] Zhi-Yong Wang[1]†

Brassinosteroid (BR) homeostasis and signaling are crucial for normal growth and development of plants. BR signaling through cell-surface receptor kinases and intracellular components leads to dephosphorylation and accumulation of the nuclear protein BZR1. How BR signaling regulates gene expression, however, remains unknown. Here we show that BZR1 is a transcriptional repressor that has a previously unknown DNA binding domain and binds directly to the promoters of feedback-regulated BR biosynthetic genes. Microarray analyses identified additional potential targets of BZR1 and illustrated, together with physiological studies, that BZR1 coordinates BR homeostasis and signaling by playing dual roles in regulating BR biosynthesis and downstream growth responses.

Brassinosteroids (BRs) are plant hormones that play essential roles in growth and development. Mutant plants that are defective in BR biosynthesis or signaling display characteristic phenotypes including dwarfism, curled leaves, male sterility, and light-grown morphology in the dark (*1–4*). Unlike many plant hormones that undergo long-distance transport and thus have separate sites of synthesis and action, BRs appear to be synthesized and function in the same tissue or even in the same cell (*5–7*). With BRs acting at the site of synthesis, cells must monitor and tightly regulate their BR biosynthesis to achieve balanced cell expansion in normal plant development (*7*). Consequently, the expression of many BR biosynthetic genes is feedback regulated by BR signaling (*8*, *9*). Such a mechanism of feedback regulation is therefore an integral part of BR action, and a detailed knowledge of this mechanism at the molecular level is essential for an integrated understanding of how plants use BRs to regulate growth and development.

Unlike animal steroid hormones, which regulate gene expression by binding to the nuclear receptor family of transcription factors,

BRs are perceived by cell-surface receptors in plants (*10*). Studies of BR-signaling mutants and their suppressors have led to the identification of a number of BR signal transduction components. These include the cell-surface receptor kinases BRI1 and BAK1, which perceive the BR signal and initiate the signal transduction cascade (*11–13*); the nuclear protein BZR1 and its homolog BZR2/BES1, which activate growth responses and are dephosphorylated and stabilized by BR signaling (*14*, *15*); the BIN2 kinase, which negatively regulates BR responses by phosphorylating BZR1 and BZR2/BES1 and targeting them for degradation by the proteasome (*15–17*); and the BSU1 phosphatase, which positively regulates BR responses by dephosphorylating and stabilizing the BZR2/BES1 protein (*18*). These studies illustrated a signaling cascade leading from the cell-surface receptors to dephosphorylation and accumulation of BZR1 and BZR2/BES1 in the nucleus. How BZR1 and BZR2/BES1 mediate BR regulation of gene expression remains a major gap in our understanding of the BR signal transduction pathway (*19*).

Studies of the *bzr1-1D* mutant suggested that BZR1 is involved in both growth promotion and feedback regulation of BR biosynthesis (*14*). The dominant *bzr1-1D* mutation increases BZR1 protein accumulation, suppresses the BR-insensitive *bri1* and *bin2* mutants, and causes insensitivity to the BR biosynthetic inhibitor brassinazole, indicating

a positive role for BZR1 in BR signaling. However, when grown in the light, the *bzr1-1D* mutant has reduced cell elongation, reduced levels of BR, and reduced expression of *CPD*, a BR biosynthetic gene that is feedback inhibited by BR signaling (*3*, *8*, *14*). This suggests that BZR1 can also have a negative effect on BR-regulated growth, perhaps by promoting feedback inhibition of BR biosynthesis. In contrast, the homologous *BZR2/BES1* gene appears to increase cell elongation under all growth conditions (*15*). The BR repression of *CPD* and several other BR biosynthetic genes was detected as early as 15 min after BR treatment (*20*), and the BR-induced BZR1 dephosphorylation and accumulation were detected within 10 min of BR treatment (*17*). These observations suggest that BZR1 might directly repress the BR biosynthetic genes.

BZR1 is a member of a small gene family that shares no appreciable sequence similarity to any known proteins (*14*). To test whether BZR1 is a DNA binding protein that directly binds to the *CPD* promoter, we performed electrophoretic mobility shift assays (EMSAs) using radio-labeled DNA fragments of the *CPD* promoter and bacterially expressed BZR1 proteins fused to the maltose binding protein (MBP). Figure 1A shows that the MBP-BZR1 (amino acids 21 to 336) fusion protein binds to the –208 to –60 region but not to the –502 to –175 region of the *CPD* promoter (translational start is +1). The N-terminal domain (BZR1N, amino acids 21 to 104) but not the C-terminal domain (BZR1C, amino acids 90 to 336) of BZR1 showed DNA binding activity (Fig. 1B), indicating that the N-terminal domain of BZR1 is the DNA binding domain. The BZR1 DNA binding domain encoded by the first exon is the most conserved region of the BZR1-related proteins and is thus named the BZR domain. BZR1 and its homologs therefore represent a previously unknown kind of DNA binding protein that appears to be plant specific.

Using chemical and deoxyribonuclease I (DNase I) footprint assays, we further localized the BZR1 binding site in the –90 to –75 region of the *CPD* promoter containing an AAACCCCCCGTGTGC sequence (Fig. 1C). To further define the optimal DNA sequence to which BZR1 binds, we performed a random binding site selection (RBSS) experiment (*21*). The BZR1-binding DNA was selected from a pool of DNA fragments containing 16-base random sequences in the middle by

[1]Department of Plant Biology, Carnegie Institution, Stanford, CA 94305, USA. [2]Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: zywang24@stanford.edu

using the MBP-BZR1N protein immobilized on amylose agarose beads. All 28 sequences identified contain the CGTG sequence (Fig. 1D). Alignment of these sequences identified CGTG(T/C)G as the optimal binding site for BZR1 (Fig. 1D), which indeed is present in the footprinted region of the *CPD* promoter (Fig. 1C). Mutating individual residues in the CGTG sequence markedly reduced competition for BZR1 binding (Fig. 1E, m7 to m10), indicating that BZR1 binding requires a CGTG sequence. Mutation of the CGTGTG element in the *CPD* promoter to AAAAAA completely abolished BZR1 binding (Fig. 1E, m13). These results demonstrate that BZR1 is a DNA binding protein that binds to the CGTG(T/C)G sequence. In support of this conclusion, we found BZR1-binding sequences in the promoters of all four feedback-regulated BR biosynthetic genes [*CPD*, *DWF4*, *ROT3*, and *BR6OX* (*20*)], indicating a general role of BZR1 in feedback regulation. Figure 1F shows that the promoter of *DWF4* contains two BZR1-binding sequences, and either of them can bind to BZR1.

To determine whether BZR1 binds to DNA in vivo, we performed chromatin immunoprecipitation (ChIP) experiments. The BZR1-CFP fusion protein expressed in transgenic plants using the BZR1 native promoter (*14*) was immunoprecipitated using an antibody to green fluorescent protein (anti-GFP). Genomic DNA fragments that coimmunoprecipitate with BZR1-CFP were analyzed by real-time quantitative polymerase chain reaction (RT-PCR). The amount of the *CPD* and *DWF4* promoter DNAs immunoprecipitated from the BZR1-CFP transgenic plants was over 27-fold higher than that precipitated from the nontransgenic control plants and about fivefold higher than the two control genes that are not responsive to BR and contain no BRRE (molybdopterin synthase sulphurylase, *CNX5*, At5g55130; and ubiquitin-conjugating enzyme, *UBC*, At5g56150) (Fig. 2A). BR treatment significantly increased the coimmunoprecipitation of the *CPD* and *DWF4* DNA but not of the control genes (Fig. 2A). These results provide direct evidence that BZR1 binds to the promoters of *CPD* and *DWF4* genes in vivo and that BR signaling increases BZR1 binding.

To determine whether BZR1 binding is required for regulation of *CPD* expression, we introduced the mutation that abolishes BZR1 binding (Fig. 1E, m13) into a *CPD* promoter–β-glucuronidase reporter gene construct (m*CPD*-GUS) and examined its effect on gene expression in transgenic *Arabidopsis* plants and in transiently transformed protoplasts. *Arabidopsis* transformed with m*CPD*-GUS expressed markedly higher GUS activity than those transformed with the wild-type *CPD*-GUS reporter gene (Fig. 2B), indicating that the BZR1 binding site is required for

repression of *CPD* expression in vivo. In protoplast transient assays, overexpression of BZR1 significantly reduced *CPD*-GUS expression, confirming that BZR1 negatively regulates *CPD* expression. By contrast, overexpression of BZR1 did not have a significant effect on the expression of the m*CPD*-GUS reporter gene (Fig. 2C). These results indicate that BZR1 binding is essential for the feedback inhibition of *CPD* gene expression. The BZR1 binding sequence [CGTG(T/C)G] is thus named BR response element (BRRE).

To determine whether BZR1 is sufficient to repress transcription, we performed transcriptional reporter gene assays using a promoter-reporter gene that contains binding sequences for the LexA and Gal4 DNA binding proteins (Fig. 2D). This reporter gene is expressed at high levels when cotransformed into protoplasts with the LexA-VP16 fusion construct, which contains the coding sequences for the LexA DNA binding domain fused in frame with the coding sequence of the VP16 transcriptional activation domain (*22*). As reported previously (*22*), the Gal4 fusion with the transcriptional repressor domain of IAA17 (IAA17α1) reduced the activation of the reporter gene by LexA-VP16 (Fig. 2D). Similarly, cotransformation of the reporter gene with a construct for overexpression of the BZR1-Gal4 fusion protein (35S-BZR1-Gal4) significantly reduced the reporter gene expression both in the absence and presence of the LexA-VP16 construct (Fig. 2D), indicating that BZR1 is a potent transcriptional repressor. Together, the above results demonstrate that BZR1 is a transcriptional repressor that, upon activation by BR signaling, binds to the BRRE of the BR biosynthetic genes and confers feedback inhibition of BR biosynthesis.
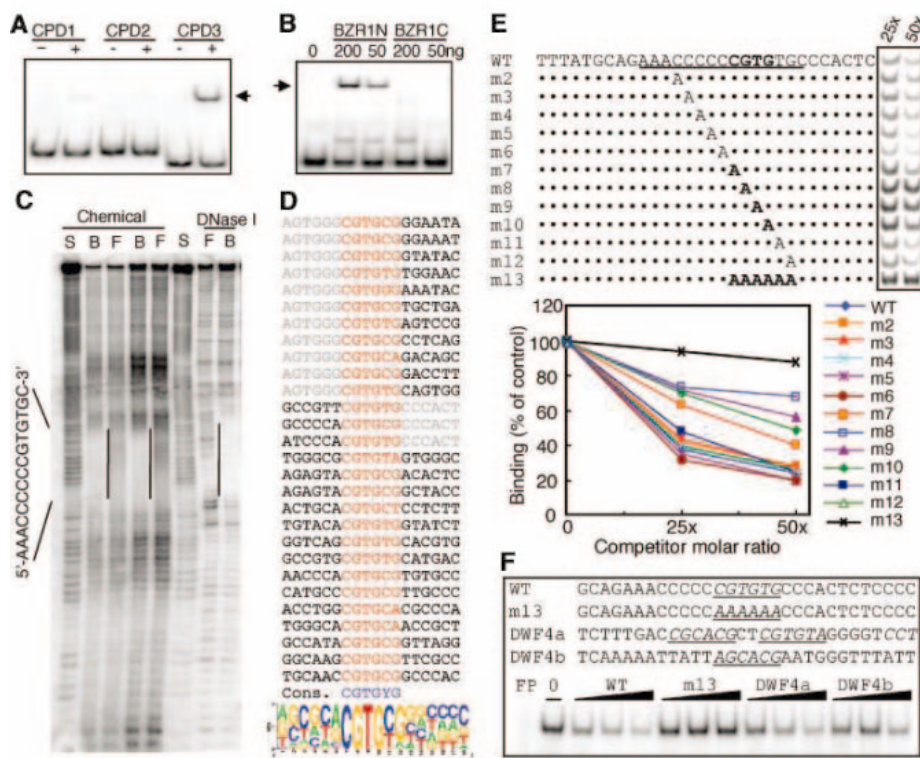
Fig. 1. BZR1 is a DNA binding protein. (**A**) EMSAs showing that the MBP-BZR1 (amino acids 21 to 336) fusion protein (BZR1) specifically binds to the –208 to –60 region (*CPD3*) but not the –502 to –339 (*CPD1*) or –348 to –175 (*CPD2*) regions of the *CPD* promoter. (**B**) MBP fusion proteins of the N-terminal domain (amino acids 22 to 104, BZR1N) but not the C-terminal domain of BZR1 protein (amino acids 90 to 336, BZR1C) binds to the *CPD* promoter. (**C**) DNA footprint assays of the BZR1 binding site in the *CPD* promoter. Free (F) and protein-bound (B) DNAs were cleaved by either phenanthroline-copper (Chemical) in gel or DNase I in solution and separated on a sequencing gel. Lane S is the G + A chemical sequencing reaction. Footprint regions are marked by vertical lines and corresponding sequence. (**D**) RBSS for optimal BZR1 binding site. Shown on the top is the alignment of all identified sequences, with the matching sequences marked in red. The nonrandom residues immediately flanking the matching sequences are marked in light gray and are excluded from calculating the consensus sequence and logo of the alignment (bottom). (**E**) Oligonucleotide sequences of the *CPD* promoter containing the wild-type (WT) or mutant (m2 to m13) BZR1 binding site were used as competitors at 25× or 50× molar ratios in EMSAs. Image and quantification of shifted bands are shown. (**F**) Competition of MBP-BZR1N binding to the BRRE of *CPD* by potential BRREs from the *DWF4* (*DWF4a* and *DWF4b*) gene. Black wedges represent increasing amount of competitors (12.5×, 25×, and 50× in molar excess). FP, free probe with no protein.

Such a direct role of BZR1 in feedback inhibition of BR biosynthetic genes can potentially explain the dwarf phenotype of light-grown *bzr1-1D* mutant plants. Whereas *bzr1-1D* mutation increases cell elongation in the dark or in the absence of upstream BR signaling, *bzr1-1D* mutant plants grown in the light show a weak dwarf phenotype with reduced hypocotyl and petiole lengths and dark green curled leaves (*14*). These light-grown *bzr1-1D* plants resemble BR-deficient and -insensitive mutants. Such opposite growth phenotypes of *bzr1-1D* in the dark and light suggest that BZR1 is regulated by light. Light either switches BZR1 from a positive regulator to a negative regulator of BR signaling or alters the contributions of BZR1 to feedback inhibition of BR biosynthesis and to growth responses (*14*). To determine whether the reduced BR level is responsible for the dwarf phenotypes of the light-grown *bzr1-1D*, we tested the effects of

both exogenous BR application and increasing endogenous BR synthesis on the dwarf phenotype of *bzr1-1D*. The hypocotyls of the *bzr1-1D* mutant seedlings are shorter than those of the wild type on regular medium but longer than those of the wild type on medium containing 10 nM brassinolide (BL, the most active brassinosteroid) (Fig. 3A), indicating that the *bzr1-1D* mutant is hypersensitive to BR. By contrast, the *bes1-D* mutant seedlings have long hypocotyls that are not affected by BL (Fig. 3A). Consistent with the gain-of-function phenotypes of *bes1-D*, the transferred DNA (T-DNA) knockout *bes1-1* mutant showed slightly reduced hypocotyl length and weaker response to BR treatment. The short hypocotyl phenotype of *bzr1-1D* was also suppressed by a transgenic line that overexpresses the BR biosynthetic gene *DWF4 (DWF4-ox)*, because the hypocotyl lengths of the *bzr1-1D/DWF4-ox* double homozygous plants

were similar to those of the *DWF4-ox* plants (Fig. 3A). These results indicate that the dwarf phenotype of light-grown *bzr1-1D* is due to BR deficiency caused by increased feedback inhibition of BR biosynthesis and that *bzr1-1D* has increased BR sensitivity.

The BR hypersensitivity of *bzr1-1D* is consistent with its partial suppression of *bri1* in the light (*14*) and supports a direct positive effect of *bzr1-1D* on downstream growth responses in the light. To determine whether BZR1 also mediates feedback regulation in the dark, we performed quantitative expression analysis of the BR biosynthetic genes *CPD*, *DWF4*, and *BR6OX* in both light and dark conditions using RT-PCR. The results indicate that the expression of these genes in *bzr1-1D* was reduced to a similar extent in the dark and in the light (Fig. 3B). Thus, BZR1 appears to mediate both growth promotion and feedback regulation of BR biosynthesis in both light and dark conditions. The opposite hypocotyl phenotypes of *bzr1-1D* in the dark and light suggest that the direct growth-promoting effect of *bzr1-1D* is sufficient to overcome the negative effect of BR deficiency in the dark but not in the light. The dwarf phenotype of light-grown *bzr1-1D* thus suggests that a BZR1-independent BR pathway contributes to growth promotion in the light and is inactivated as a result of BR deficiency in *bzr1-1D*.

To identify additional BZR1-regulated genes and to understand the BR-regulated transcriptional pathways, we examined the effects of *bzr1-1D* and *det2* mutations on the expression of BR-regulated genes by using the *Arabidopsis* full-genome oligo microarray (Affymetrix). Based on previous studies and public microarray data (*20*, *23*, *24*), 324 early BR-responsive genes (greater than twofold change within 3 hours of BR treatment, full-genome array) and 235 late BR-responsive genes (greater than twofold change at 12 or 24 hours but not at 3 hours of BR treatment, 8300-gene array) (*23*) were selected for our analysis. These genes were further classified into BR-induced and BR-repressed genes, and those genes affected by *bzr1-1D* (i.e., that differ from the wild type by more than 20%, *P* < 0.2) were analyzed in detail (Fig. 3C). Unlike *det2*, which consistently affects most genes in the opposite way to that of BR treatment, *bzr1-1D* appears to have different effects on different sets of BR-regulated genes (Table 1 and Fig. 3C). First, expression of nearly all the *bzr1-1D*–affected, early BR-repressed genes was reduced in *bzr1-1D*, which is consistent with BZR1 directly regulating early BR-repressed genes. Second, more of the late BR-induced genes were down-regulated than up-regulated (72% versus 28%) in *bzr1-1D*, which could explain the dwarf phenotype of light-grown *bzr1-1D*. For the early BR-induced and late BR-repressed
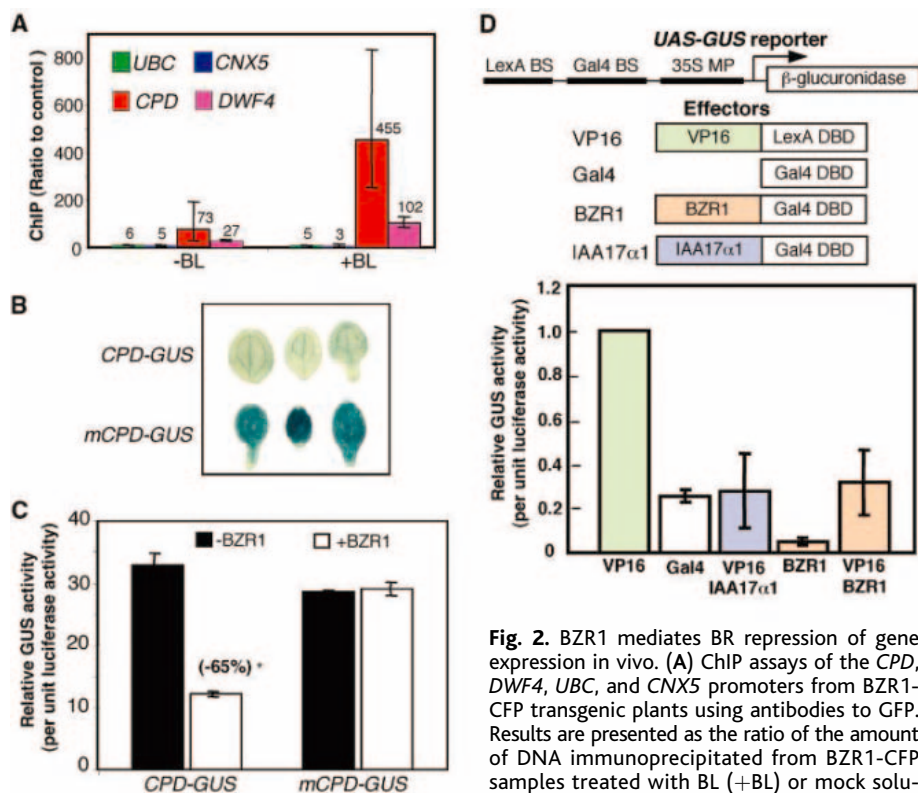
Fig. 2. BZR1 mediates BR repression of gene expression in vivo. (**A**) ChIP assays of the *CPD*, *DWF4*, *UBC*, and *CNX5* promoters from BZR1-CFP transgenic plants using antibodies to GFP. Results are presented as the ratio of the amount of DNA immunoprecipitated from BZR1-CFP samples treated with BL (+BL) or mock solution (−BL) to that from nontransgenic control plants, determined by quantitative RT-PCR. The results (mean ± SE) from six PCR analyses of two biological repeats are shown. (**B**) Expression of the *CPD*-GUS and *mCPD*-GUS reporter genes in *bzr1-1D* plants. Cotyledons of 10-day-old seedlings from three independent T1 transgenic lines were histochemically stained for GUS enzyme activity. (**C**) Expression of *CPD*-GUS or *mCPD*-GUS reporter gene in *Arabidopsis* protoplasts transformed without (−) or together with (+) the BZR1 overexpression construct (*35S*-BZR1), relative to the *35S*-luciferase internal control. (**D**) Transient assays of transcriptional activity of BZR1. Protoplasts from a tobacco BY-2 cell line were transformed with the reporter (*UAS*-GUS) and effector constructs (upper panel), and the reporter gene expression is shown in the bottom panel. *UAS*-GUS, reporter construct containing Gal4 and LexA binding sites and a *35S* minimal promoter upstream of the coding sequence of GUS; VP16, VP16 fused to the LexA DNA binding domain (DBD); Gal4, Gal4 DBD; IAA17α1, the transcription repression domain of IAA17 fused to the Gal4 DBD; BZR1, full-length BZR1 fused to Gal4 DBD. *35S*-luciferase was cotransformed as internal control, and the GUS reporter gene expression was normalized to the luciferase activity and presented as values relative to the VP16 control. Error bars indicate SE.

genes, the number of genes that were activated was similar to the number that were reduced in *bzr1-1D*. In agreement with physiological studies (Fig. 3A), these results provide evidence that *bzr1-1D* has positive effects on BZR1 downstream components and negative effects on BZR1-independent BR pathways by reducing BR levels. Genes affected similarly by *bzr1-1D* and *det2* but oppositely by BR treatment are likely regulated by the BZR1-independent pathways that are negatively affected by BR deficiency in *bzr1-1D*. Conversely, genes affected similarly by *bzr1-1D* and BR treatment are likely downstream targets of BZR1. Of those genes, the early BR-repressed are potentially direct targets of BZR1, and the late BR-responsive genes are most likely indirectly regulated by BZR1. The increased expression of 17 early BR-induced genes in *bzr1-1D* suggests that BZR1 might also function as a transcriptional activator; however, genes that respond within 3 hours of BR treatment could also be indirectly regulated by BZR1.

The presence of the BRRE in the promoters of various BR-regulated genes was analyzed. Consistent with BZR1 being a BR-activated transcriptional repressor, the BRRE was overrepresented in the proximal promoter regions of the early BR-repressed genes (Fig. 3D), as compared with all genes of the *Arabidopsis* genome or with the BR-induced genes. Most of the BR-repressed genes containing BZR1 binding sites were also repressed in the *bzr1-1D* mutant, further supporting the conclusion that they are likely direct target genes of BZR1 (table S1). These genes include the BR biosynthetic genes *CPD*, *DWF4*, *ROT3*, and *BR6OX* (*20*).

Although *bzr1-1D* mutation has positive effects on many BR-induced genes (Fig. 3B), the finding that the BZR1 binding-site sequence is not greatly enriched in BR-induced genes suggests that BZR1 does not directly activate BR-induced genes. However, it is possible that BZR1 activates gene expression by interacting with other DNA binding proteins that bind directly to BR-induced promoters. It is more likely that BZR1 indirectly activates gene expression by repressing the expression of other transcriptional repressors. Previous microarray studies have shown that most BR-induced genes respond to BR treatment very slowly as compared with BR-repressed or auxin-induced genes (*20*, *23*), suggesting that BR induces these genes indirectly through transcriptional relay. This observation is consistent with auxin and BR regulating gene expression by promoting (*25*) and preventing the degradation of transcriptional repressors, respectively. Potential transcriptional repressors that are repressed by BZR1 include two Myb-related DNA binding proteins (At4g01680 and At1g71030) and the BZR1 homolog BZR3

(At4g36780) (table S1). These genes contain the BRRE in their promoters and are repressed by BR and *bzr1-1D*. They are therefore potential direct targets of BZR1. Functional studies of these BZR1-regulated transcription factors will further illustrate the circuit for BR regulation of gene expression.

This study illustrates the molecular mechanisms by which BR signaling regulates gene expression and defines BZR1 as a central regulator for coordinating BR signaling, BR biosynthesis, and growth responses. We demonstrate that BR-regulated gene expression is mediated directly by the DNA binding protein BZR1 and, likely, its homologs, which are stabilized by upstream BR signaling. Together with previous studies, our results support a model for the function of BZR1 in the BR signal transduction pathway as diagrammed in fig. S1. When BR level is low, BZR1 is phosphorylated by
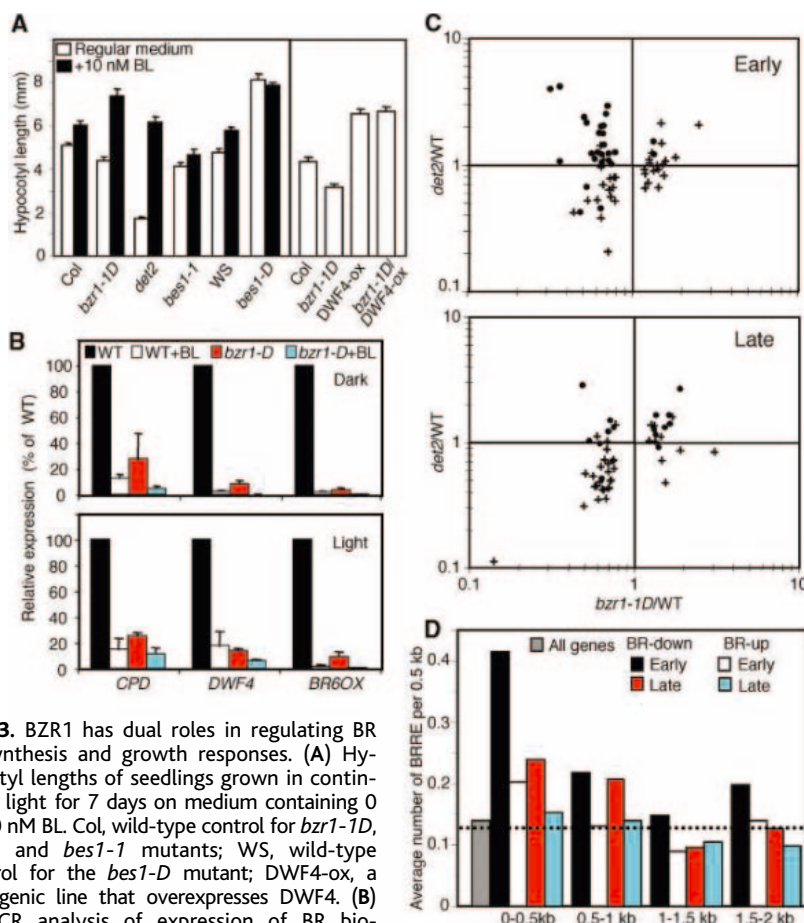
Fig. 3. BZR1 has dual roles in regulating BR biosynthesis and growth responses. (**A**) Hypocotyl lengths of seedlings grown in continuous light for 7 days on medium containing 0 or 10 nM BL. Col, wild-type control for *bzr1-1D*, *det2*, and *bes1-1* mutants; WS, wild-type control for the *bes1-D* mutant; DWF4-ox, a transgenic line that overexpresses DWF4. (**B**) RT-PCR analysis of expression of BR biosynthetic genes in the *bzr1-1D* mutant and wild-type (WT) seedlings grown in the dark (upper) or light (lower panel) and after 3 hours of BL treatment (+BL). The results (mean ± SD) of three experiments are shown. (**C**) Microarray analysis of BR-regulated genes in *det2*, *bzr1-1D*, and wild-type (WT) *Arabidopsis*. *y* axis, ratio of expression between *det2* and wild type; *x* axis, ratio of expression between *bzr1-1D* and wild type. (Upper panel) Early BR-responsive genes; (lower panel) late BR-responsive genes. Dots indicate BR-induced genes and crosses indicate BR-repressed genes. (**D**) Average number of BRREs in the promoter regions of genes that are up- (BR-up) or down- (BR-down) and early- or late-regulated by BR treatment. Dashed line, calculated probability of occurrence of the BRRE sequence in any 500–base pair *Arabidopsis* intergenic sequence. Gray bar, average number of BRREs in the 500–base pair upstream regions of all genes in the *Arabidopsis* genome.

Table 1. Summary of microarray analysis of the *bzr1-1D* mutant. Numbers of genes regulated by BR treatment and altered in the *bzr1-1D* mutant are shown.

BR response	BR-regulated genes	Altered in *bzr1-1D*	Reduced in *bzr1-1D*	Increased in *bzr1-1D*
Early down	101	26 (25.7%)	24 (92.3%)	2 (7.7%)
Early up	223	33 (14.8%)	16 (48.5%)	17 (51.5%)
Late down	63	17 (26.7%)	8 (47.1%)	9 (52.9%)
Late up	172	36 (20.9%)	26 (72.2%)	10 (27.8%)

BIN2 and then degraded by the proteasome (*17*), and the lack of BZR1 would in turn allow expression of BR biosynthetic genes and increase the BR level. When BR level is high, BR activation of BRI1 and BAK1 receptor kinases leads to dephosphorylation of BZR1 by either inhibiting BIN2 kinase or activating BSU1 phosphatase or both (*12*, *18*, *19*). Dephosphorylation of BZR1 prevents its degradation by the proteasome and increases its accumulation in the nucleus (*17*). BZR1 then binds to the promoters and inhibits the expression of BR biosynthetic genes to reduce the BR level. BZR1 also promotes growth responses by regulating downstream BR-responsive genes. Such dual roles of BZR1 in regulating BR biosynthesis and growth responses ensure optimal levels of BR action for plant development, making BZR1 a central regulator of the BR pathway. The BZR1 feedback loop is thus the key mechanism for maintaining the equilibrium of BR action. How plant cells reset this equilibrium in response to developmental and environmental signals is a question to be answered in future studies.

References and Notes

1. S. D. Clouse, M. Langford, T. C. McMorris, *Plant Physiol.* **111**, 671 (1996).
2. T. Altmann, *Trends Genet.* **14**, 490 (1998).
3. M. Szekeres et al., *Cell* **85**, 171 (1996).
4. J. Li, K. H. Nam, D. Vafeados, J. Chory, *Plant Physiol.* **127**, 14 (2001).
5. G. J. Bishop, K. Harrison, J. D. Jones, *Plant Cell* **8**, 959 (1996).
6. Y. Shimada et al., *Plant Physiol.* **131**, 287 (2003).
7. G. M. Symons, J. B. Reid, *Plant Physiol.* **135**, 2196 (2004).
8. J. Mathur et al., *Plant J.* **14**, 593 (1998).
9. S. Bancos et al., *Plant Physiol.* **130**, 504 (2002).
10. C. S. Thummel, J. Chory, *Genes Dev.* **16**, 3113 (2002).
11. J. Li et al., *Cell* **110**, 213 (2002).
12. J. Li, *Curr. Opin. Plant Biol.* **6**, 494 (2003).
13. Z. Y. Wang, H. Seto, S. Fujioka, S. Yoshida, J. Chory, *Nature* **410**, 380 (2001).
14. Z. Y. Wang et al., *Dev. Cell* **2**, 505 (2002).
15. Y. Yin et al., *Cell* **109**, 181 (2002).
16. J. Li, K. H. Nam, *Science* **295**, 1299 (2002).
17. J. X. He, J. M. Gendron, Y. Yang, J. Li, Z. Y. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10185 (2002).
18. S. Mora-Garcia et al., *Genes Dev.* **18**, 448 (2004).
19. Z. Y. Wang, J. X. He, *Trends Plant Sci.* **9**, 91 (2004).
20. H. Goda, Y. Shimada, T. Asami, S. Fujioka, S. Yoshida, *Plant Physiol.* **130**, 1319 (2002).
21. T. K. Blackwell, H. Weintraub, *Science* **250**, 1104 (1990).
22. S. B. Tiwari, G. Hagen, T. J. Guilfoyle, *Plant Cell* **16**, 533 (2004).
23. H. Goda et al., *Plant Physiol.* **134**, 1555 (2004).
24. H. Goda, S. Yoshida, Y. Shimada, The *Arabidopsis* Information Resource (TAIR), available at http://arabidopsis.org/servlets/TairObject?type=hyb_descr_collection&id=1007966053.
25. N. Dharmasiri, M. Estelle, *Trends Plant Sci.* **9**, 302 (2004).
26. Molecular interaction data have been deposited in the Biomolecular Interaction Network Database (BIND) with accession codes 197445 and 197446. We thank M. Szekeres for providing the *CDP*-GUS constructs; T. J. Guilfoyle for the *UAS*-GUS, *VP16*-*LexA*, and *IAA17α1-Gal4* constructs; V. Walbot for the *35S*-Luciferase vector; W. Frommer for the BY2 cells; B-H. Hou and T. Hamman for assistance with microarray analysis; Y. Yang and N. Marinova for technical assistance; Y. Lou for assistance with promoter cis-element analysis; and W. Briggs, D. Bergmann, D. Ehrhardt, Z. He, A. Grossman, and C. Somerville for helpful comments on the manuscript. This work was supported by grants from NIH (R01 GM66258-01), the U.S. Department of Energy (DE-FG02-04ER15525), and the Carnegie Institution of Washington (Z.-Y.W.), and by a training grant from NIH (5T32GM007276) (J.M.G.).

Insect Sex-Pheromone Signals Mediated by Specific Combinations of Olfactory Receptors

Takao Nakagawa,[1] Takeshi Sakurai,[2,4] Takaaki Nishioka,[3,4]* Kazushige Touhara[1]*

We describe two male-specific olfactory receptors (ORs) in the silk moth, *Bombyx mori*, that are mutually exclusively expressed in a pair of adjacent pheromone-sensitive neurons of male antennae: One is specifically tuned to bombykol, the sex pheromone, and the other to bombykal, its oxidized form. Both pheromone ORs are coexpressed with an OR from the highly conserved insect OR subfamily. This coexpression promotes the functional expression of pheromone receptors and confers ligand-stimulated nonselective cation channel activity. The same effects were also observed for general ORs. Both odorant and pheromone signaling pathways are mediated by means of a common mechanism in insects.

There are two distinct chemical perception mechanisms in the antennae of the insect olfactory system (*1*). The first perception mechanism, the "generalist" system, recognizes odorants from foods and plants. The olfactory receptor (OR) family, which belongs to the seven-transmembrane G protein–coupled receptor family and includes about 60 multigenes in insects, is responsible for the first step of chemosensation in the olfactory neurons of antennae (*2–4*). Studies of olfaction in fruit flies have shown that odorants are discriminated by a combinatorial receptor code mediated by broadly tuned ORs (*5*).

The other chemical perception mechanism is the "specialist" system, which detects pheromones that are elicited by conspecifics (*6*). Sex pheromones released by adult female insects are detected by narrowly tuned olfactory neurons in the conspecific male antennae. Notably, insects can discriminate a subtle difference in stereochemistry or chirality of molecules, and they also use the specific ratios of two or a few components as a species-specific cue (*7*). The silk moth, *B. mori*, possesses the simplest pheromone communication system, wherein one achiral compound, bombykol, elicits the full array of sexual behaviors (*8*).

We have recently cloned a male-specific OR gene, *BmOR1*, that encodes a sex-pheromone receptor in *B. mori* (*9*). Ectopic expression of BmOR1 in female antennae conferred electrophysiological responses to bombykol, providing evidence that BmOR1 is a bombykol receptor (*9*). The specific response to bombykol was also seen in *Xenopus* oocytes expressing BmOR1 and the G protein BmGαq. However, only 10 to 15% of the injected oocytes were responsive. Furthermore, high concentrations of bombykol were necessary to activate BmOR1. These observations led us to speculate that additional components are necessary to fully reconstitute the pheromone signaling pathway in a heterologous cell system.

Members of a highly conserved OR gene subfamily are expressed in many of the receptor neurons in the antennae of insects (*4*, *10*, *11*). OR neurons in *Or83b* mutant flies showed no odorant-evoked action potential and little spontaneous activity (*12*). Thus, the Or83b family does not seem to play a role in direct detection of odorants or pheromones but rather acts as a dimerization partner for pheromone ORs. BmOR2 in *B. mori* appears to be a member of this family of conserved ORs (*9*, *13*). We thus examined

[1]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8562, Japan. [2]Division of Applied Biosciences and [3]Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan. [4]Core Research for Evolution Science and Technology of Japan Science and Technology Agency, Kawaguchi, 332–0012 Japan.

*To whom correspondence should be addressed. E-mail: touhara@k.u-tokyo.ac.jp (K.T.); nishioka@scl.kyoto-u.ac.jp (T.N.)

Fig. 1. Coexpression of BmOR1 and BmOR2 and electronic responses to bombykol. (**A**) Two-color fluorescent in situ hybridizations of 15-μm sections from a male antenna. BmOR1 [Digoxigenin-RNA (DIG-RNA)] is shown in green and BmOR2 (fluorescein RNA) in magenta. Scale bar, 20 μm. Out of 369 neurons expressing BmOR2 in 21 sensillum sections from five antennae, 159 neurons expressed BmOR1. (**B**) Bombykol-induced currents in *Xenopus* oocytes. (Top) Current traces of oocytes injected with the indicated cRNAs in response to 30 μM bombykol with depolarization step pulses. (Bottom) Current traces recorded at –80 mV. Lysophosphatidic acid (LPA) (1 μM) was used as a control for the induction of Ca^{2+}-dependent Cl^- current in oocytes. (**C**) (Top) Current trace with application of the indicated concentrations of bombykol. Bombykol

was sequentially applied to the same oocytes. (Bottom) The corresponding dose-response curve. Each point represents the mean current value (±SEM) from four independent oocytes. (**D**) Normal bath solution showed a reversal potential of –7.0 ± 2.0 mV ($n = 5$). For ion exchange experiments, currents obtained before bombykol application were digitally subtracted from the curves. Na^+-free solution caused a negative shift of the reversal potential of –54.1 ± 1.1 mV ($n = 4$). High-K^+ solution did not cause a significant shift in the reversal potential [–8.9 ± 1.9 mV ($n = 5$)]. High-Ba^{2+} solution resulted in a slight shift in the reversal potential to –50.1 ± 1.3 mV ($n = 5$). Error bars show mean ±SEM. (**E**) Effects of channel blockers on bombykol-induced currents. (Top) The current trace was recorded by applying bath solution containing 2-APB (50 μM) or RR (50 μM). (Bottom) Quantitative analysis of the inhibition of bombykol-induced currents by channel blockers. Data represent the means (±SEM) from five independent oocytes. [(B), (C), and (E)] Ligands were applied for 10 s at the time indicated by the arrowhead.
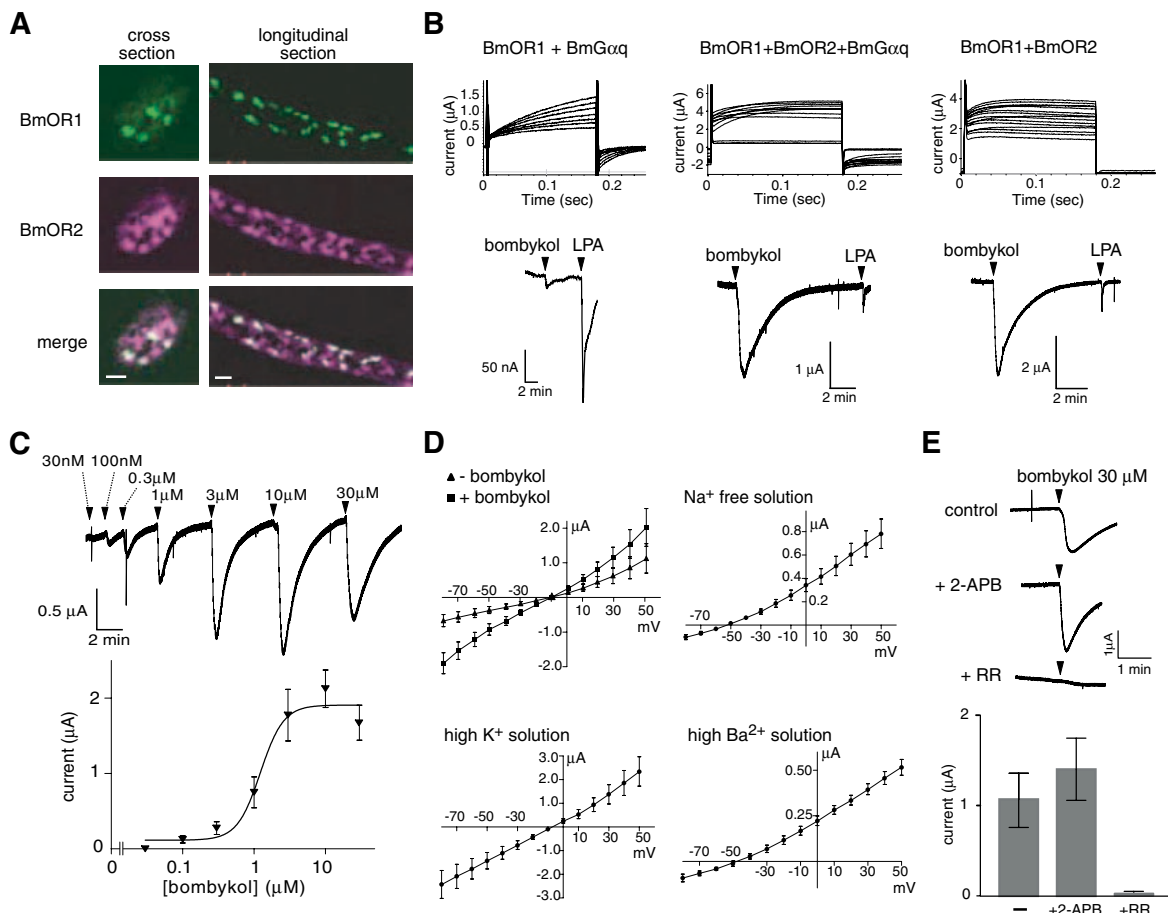
the expression patterns of BmOR1 and BmOR2 in the male adult antenna by two-color double in situ hybridization. BmOR2 was broadly expressed in antennal neurons, and 43% of BmOR2-labeled cells coexpressed BmOR1 (Fig. 1A). Furthermore, all of the BmOR1-positive neurons expressed BmOR2, suggesting that these two ORs act together in sex pheromone detection in vivo. In addition, the level of BmOR1 was consistently higher in the membrane fraction of *Xenopus* oocytes coexpressing BmOR2 (fig. S1). Other members of the highly conserved family, including HR2, also enhanced the level of BmOR1 in the membrane fraction, whereas conventional ORs, such as HR6 (*4*), had no effect. These results suggest that BmOR2 ensures functional expression of BmOR1 by acting as a chaperone or an accessory protein.

We next conducted two-electrode voltage clamp recording in *Xenopus* oocytes coinjected with complementary RNAs (cRNAs) encoding BmOR1, BmOR2, and BmGαq. Coexpres-

sion of BmOR2 increased the percentage of bombykol-responsive oocytes to more than 95% and induced larger currents than those in oocytes lacking BmOR2 (Fig. 1B). In the absence of BmOR2, the response amplitude (±SEM) at a holding potential of –80 mV and with 30 μM bombykol was 22.8 ± 3.2 nA ($n = 6$), whereas in the presence of BmOR2 and with 30 μM bombykol, the response amplitude was 2.1 ± 0.26 μA ($n = 8$) (Fig. 1B). Surprisingly, oocytes expressing BmOR1 and BmOR2 without BmGαq also showed a response to bombykol (Fig. 1B). This outward current with depolarizing pulses, however, was different from the slow activating Cl^- current that is due to BmGαq-mediated Ca^{2+} increases. Bombykol-induced increases in currents were dose dependent, with a 50% effective concentration EC_{50} of 1.5 μM. This is one order of magnitude smaller than that previously obtained in oocytes lacking BmOR2 coexpression (*9*) (Fig. 1C). Finally, the threshold concentration was approximately 100 nM.

BmOR2 thus greatly enhances the sensitivity of BmOR1 to bombykol and initiates a previously unrecognized signaling cascade.

We next characterized the electrophysiological properties of bombykol responses in oocytes expressing BmOR1 and BmOR2 by current-voltage analysis. The current-voltage relationship of the bombykol-activated conductance was nearly linear with a reversal potential of –7.0 ± 2.0 mV ($n = 5$) (Fig. 1D). This channel property was different from that of the Ca^{2+}-activated Cl^- channel in oocytes and thus indicated that a nonselective cation channel was involved in the response to bombykol. Current-voltage relationships with various extracellular ion compositions showed that the bombykol-induced current was carried preferentially by monovalent cations (Fig. 1D). The current induced by bombykol was substantially inhibited by ruthenium red (RR), a blocker of the inositol-1,4,5-triphosphate–dependent channel and some transient receptor potential (TRP) channels, but not by 2-

aminoethoxydiphenylborane (2-APB), which is an inhibitor for other TRP channels (*14*) (Fig. 1E).

We next examined whether other members of the Or83b family behaved similarly.

The same nonselective cation channel activity in response to bombykol was observed when BmOR1 was coexpressed with HR2 or Or83b (87 and 62% amino acid sequence identity with BmOR2, respectively) but not

upon coexpression with a conventional OR such as HR6 (Fig. 2A). To further determine whether the Or83b family plays the same role in general odorant responses, we assayed the responses in oocytes coexpressing Or83b and Or47a, a *Drosophila* OR whose ligand has been identified by in vivo expression (*5*). Oocytes expressing Or47a alone or Or83b alone showed no response to any of the cognate ligands for Or47a (Fig. 2B and fig. S2), whereas oocytes expressing both Or47a and Or83b responded to pentyl acetate and weakly to 2-heptanone but not to methyl salicylate or ethyl acetate (Fig. 2B). These results demonstrate that Or47a is a general OR, the ligand specificity of which appeared to be consistent with that of Or47a-expressing neurons (*5*). Thus, the Or83b family plays a common functional role in both odorant and pheromone signaling pathways, conferring nonselective cation channel activity, by interacting with conventional ORs (*15*). Oocytes coexpressing BmOR2 and the mouse eugenol receptor (mOR-EG) or the β2-adrenergic receptor did not respond to eugenol or isoproterenol, respectively, suggesting that the function of the Or83b family is limited to the insect ORs (Fig. 2C).

We identified an additional 29 putative ORs in *B. mori* from the Silkworm Genome database, and we found that two of them, BmOR3 and BmOR4, were expressed specifically in the male antenna and that two
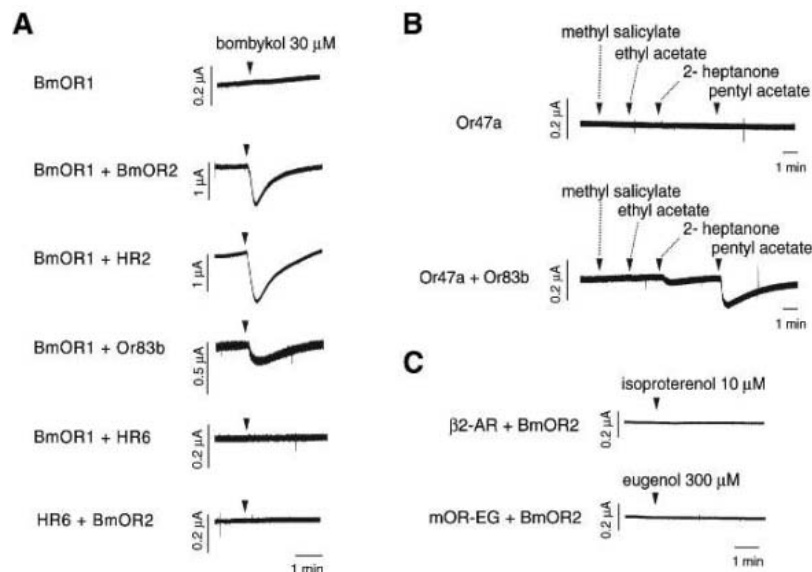
Fig. 2. Effects of coexpression of the Or83b family on the pheromone or odorant responses in oocytes. Representative inward current recorded from oocytes injected with indicated cRNAs at a holding potential of –80 mV. A pheromone or an odorant was applied for 10 s at the time indicated by the arrowhead: 30 μM bombykol (**A**); 1 mM methyl salicylate, ethyl acetate, 2-heptanone, and pentyl acetate (**B**); or 10 μM isoproterenol for β2-adrenergic receptor (β2-AR) and 300 μM eugenol for mOR-EG (**C**).
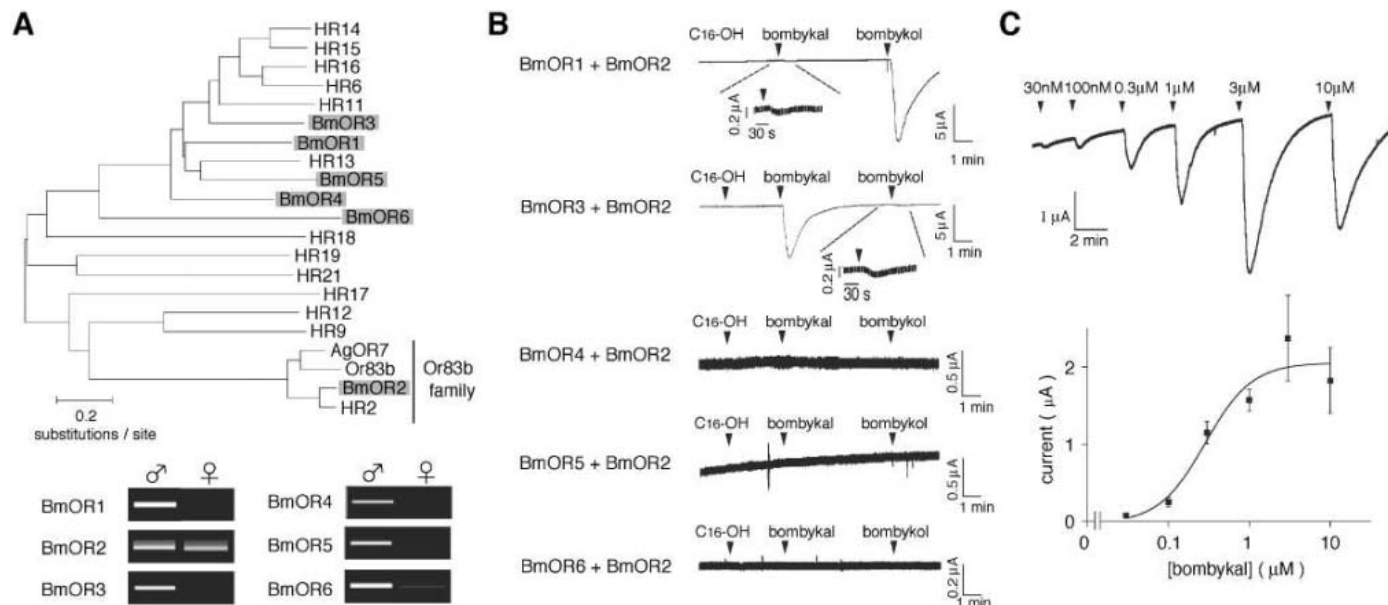
Fig. 3. Characterization of male-specific BmOR genes in *B. mori* and identification of a bombykal receptor. (**A**) (Top) Phylogenic tree of moth ORs. The phylogenic tree was generated from an alignment of the entire amino acid sequences of moth ORs with the MEGA-2 program (Molecular Evolutionary Genetics Analysis v2.1, www.megasoftware.net). HR, *Heliothis virescens* OR (*24*); AgOR, *Anopheles gambiae* OR. (Bottom) Sex-specific expression pattern of BmOR genes in an antenna as determined by reverse transcription polymerase chain reaction. (**B**) Ligand specificity of BmORs. The current trace was recorded at –80 mV with sequential application of 30 μM hexadecanol (C_{16}-OH),

bombykal, and bombykol to the same oocyte expressing the indicated set of cRNAs. The chemicals were applied for 10 s at the time indicated by the arrowhead. The insets show a higher magnification of the current trace. (**C**) Dose-dependent responses to bombykal in oocytes expressing BmOR3 and BmOR2. (Top) Current trace recorded at –80 mV by application of the indicated concentration of bombykal. Bombykal was sequentially applied for 10 s to the same oocytes at the time indicated by an arrowhead. (Bottom) Dose-response curve of oocytes injected with BmOR3 and BmOR2. Each point represents the mean current value (±SEM) from four independent oocytes.

receptors, BmOR5 and BmOR6, were dominantly expressed in males (Fig. 3A). Bombykal, an oxidized form of bombykol, another compound that is released from the female pheromone gland, elicits responses only in male antennae and has an inhibitory effect on bombykol-induced wing vibration and movement of male moths (16). We therefore tested both bombykol and bombykal in addition to hexadecanol (C_{16}-OH) as a negative control to characterize the responses of these male-specific or male-dominant ORs. BmOR1-BmOR2 combination conferred a response to bombykol [10.0 ± 2.27 µA at 30 µM ($n = 4$)], a very weak response to bombykal [0.23 ± 0.05 µA at 30 µM ($n = 4$)], and no response to C_{16}-OH (Fig. 3B). The BmOR3-BmOR2 combination elicited a response to bombykal [6.8 ± 0.15 µA at 30 µM ($n = 4$)], a very weak response to bombykol [0.14 ± 0.02 µA at 30 µM ($n = 4$)], and no response to C_{16}-OH (Fig. 3B). This effect of bombykal was dose dependent with an EC_{50} value of 0.26 µM (Fig. 3C), and the threshold concentration for activation was approximately 30 nM. The channel properties of the BmOR1-BmOR2 and BmOR3-BmOR2 combinations were similar, suggesting that the same signal

transduction mechanism was involved for both. Neither bombykol nor bombykal activated BmOR4, BmOR5, or BmOR6. Furthermore, 41 odorants that had previously been reported to elicit responses in *B. mori* antennae (17) activated neither BmOR1 nor BmOR3 (18), suggesting that these two ORs are specialist pheromone receptors that possess a high degree of specificity. Ligands for BmOR4, BmOR5, and BmOR6 remain to be identified.

In situ hybridization showed that BmOR3 was colocalized with BmOR2 in olfactory neurons in the antennae of male moths (Fig. 4A). Of the BmOR2-positive cells, 48% expressed BmOR3. BmOR1 and BmOR3 were not colocalized and were mutually exclusively expressed in two adjacent olfactory neurons (Fig. 4A). The expression of BmOR1 and BmOR3 was localized to the olfactory neurons of a trichodea sensillum. This was confirmed by double in situ hybridization with pheromone-binding protein (PBP), which is expressed in the supporting cells that surround pheromone-sensitive neurons in the male moth antenna (19) (Fig. 4B). Our results are consistent with previous physiological studies in which one of a pair of

pheromone-sensitive neurons in a long sensillum trichodeum was activated by bombykol and the other responded to bombykal (16). PBP is involved in pheromone detection and possibly pheromone discrimination (20). Thus, PBP may play a role in dissolving pheromones in the sensillum lymph, thereby lowering the threshold concentration necessary for BmOR1 activation.

Specialist neurons express a single type of male-specific OR that appears to possess a narrow specificity for pheromones: BmOR1 for bombykol and BmOR3 for bombykal. This mutually exclusive expression pattern of two pheromone receptors in single sensillum may provide a common paradigm to ensure detection of the specific ratios of a pheromone blend in the moth antenna. The neurons expressing pheromone receptors are projected into the macroglomerular complex in the male brain where pheromone information is further integrated to elicit the behavioral movement (21, 22). In contrast, generalist neurons express other ORs that detect general odorants (5), and the combinatorial code for each odorant is transmitted to distinct areas in the brain (23). Both specialist and generalist neurons require coexpression of the Or83b family, which appears to promiscuously couple with conventional ORs and ensure their chemoreceptor function by mediating nonselective cation channel activity. Although clarification of the in vivo mechanisms underlying OR-mediated current generation awaits further molecular and physiological analysis, this atypical function of a seven-transmembrane receptor suggests that there are unique and previously unappreciated aspects of receptor signal transduction.

Fig. 4. Expression pattern of BmOR1 and BmOR3 in an antenna of an adult male moth. (**A**) Two-color fluorescent in situ hybridization of 15-µm sections of an antenna. Scale bar, 20 µm. (Top) Localization of BmOR2 (fluorescein RNA, green) and BmOR3 (DIG-RNA, magenta). Out of 187 neurons expressing BmOR2 in eight sensillum sections from five antennae, 90 neurons expressed BmOR3. (Middle) Localization of BmOR1 (fluorescein RNA, green) and BmOR3 (DIG-RNA, magenta). (Bottom) Higher magnification images of the boxed regions in the middle panels. (**B**) Schematic drawing of olfactory neurons and supporting cells (SC) in a pheromone-sensitive sensillum (SL). The expression patterns of BmORs and PBP are indicated by the following colors: BmOR1, magenta; BmOR3, red; PBP, green. CU, cuticle. Two-color fluorescent in situ hybridization of PBP (fluorescein RNA, green) and either BmOR1 (DIG-RNA, magenta) or BmOR3 (DIG-RNA, red). Scale bar, 5 µm. (**C**) Model of OR-mediated signal transduction of odorant or pheromone responses in insects. Each neuron expresses one type of conventional OR that is coexpressed with a member of the Or83b family. Conventional ORs bind their cognate ligands (pheromone or odorant), resulting in stimulation of a G protein–mediated pathway and activation of a nonselective channel by coupling with the Or83b family. Transmembrane topology was predicted by the TMHMM program (TransMembrane Hidden Markow Model, www.cbs.dtu.dk/services/TMHMM).
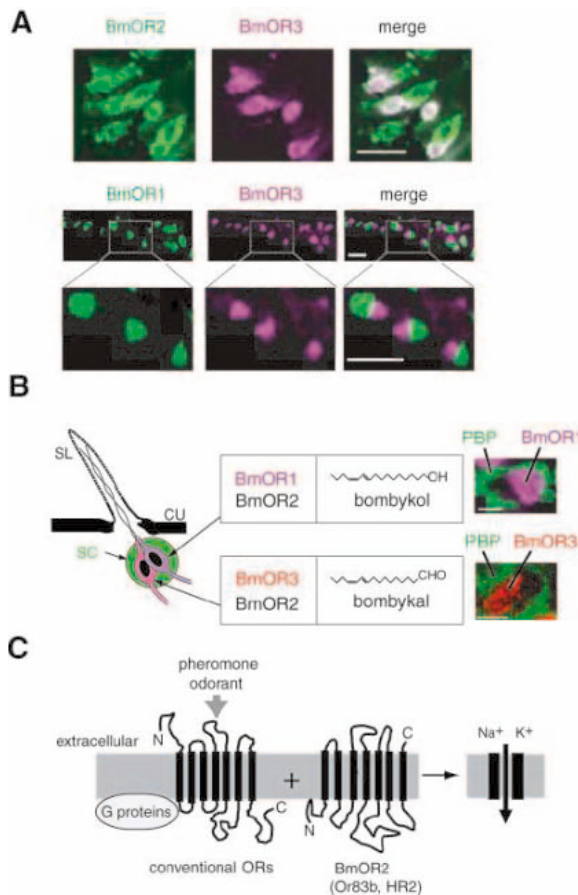
References and Notes
1. J. G. Hildebrand, G. M. Shepherd, *Annu. Rev. Neurosci.* **20**, 595 (1997).
2. L. B. Vosshall, *Chem. Senses* **26**, 207 (2001).
3. E. A. Hallem, J. R. Carlson, *Trends Genet.* **20**, 453 (2004).
4. J. Krieger et al., *Eur. J. Neurosci.* **16**, 619 (2002).
5. E. A. Hallem, M. G. Ho, J. R. Carlson, *Cell* **117**, 965 (2004).
6. K. E. Kaissling, *Chem. Senses* **21**, 257 (1996).
7. K. Mori, *Chirality* **10**, 578 (1998).
8. V. A. Butenandt, R. Beckmann, D. Stamm, E. Hecker, *Z. Naturforschg.* **14b**, 283 (1959).
9. T. Sakurai et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16653 (2004).
10. L. B. Vosshall, A. M. Wong, R. Axel, *Cell* **102**, 147 (2000).
11. R. J. Pitts, A. N. Fox, L. J. Zwiebel, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5058 (2004).
12. M. C. Larsson et al., *Neuron* **43**, 703 (2004).
13. J. Krieger, O. Klink, C. Mohl, K. Raming, H. Breer, *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **189**, 519 (2003).
14. A. Tozzi et al., *Eur. J. Neurosci.* **18**, 2133 (2003).
15. E. M. Neuhaus et al., *Nature Neurosci.* **8**, 15 (2005); published online 12 December 2004 (10.1038/nn1371).
16. K. E. Kaissling, G. Kasang, H. Bestmann, W. Stransky, O. Vostrowsky, *Naturwissenschaften* **65**, 382 (1978).
17. A. Topazzini, M. Mazza, P. Pelosi, *J. Insect Physiol.* **36**, 619 (1990).

18. T. Nakagawa, K. Touhara, unpublished data.
19. R. A. Steinbrecht, M. Laue, G. Ziegelberger, *Cell Tissue Res.* **282**, 203 (1995).
20. B. Pophof, *Chem. Senses* **29**, 117 (2004).
21. B. S. Hansson, H. Ljungberg, E. Hallberg, C. Lofstedt, *Science* **256**, 1313 (1992).
22. R. Kanzaki, K. Soo, Y. Seki, S. Wada, *Chem. Senses* **28**, 113 (2003).
23. J. W. Wang, A. M. Wong, J. Flores, L. B. Vosshall, R. Axel, *Cell* **112**, 271 (2003).
24. J. Krieger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11845 (2004).

Adaptive Coding of Reward Value by Dopamine Neurons

Philippe N. Tobler, Christopher D. Fiorillo,* Wolfram Schultz†

It is important for animals to estimate the value of rewards as accurately as possible. Because the number of potential reward values is very large, it is necessary that the brain's limited resources be allocated so as to discriminate better among more likely reward outcomes at the expense of less likely outcomes. We found that midbrain dopamine neurons rapidly adapted to the information provided by reward-predicting stimuli. Responses shifted relative to the expected reward value, and the gain adjusted to the variance of reward value. In this way, dopamine neurons maintained their reward sensitivity over a large range of reward values.

In order to select the action associated with the largest reward, it is critical that the neural representation of reward has minimal uncertainty. A fundamental difficulty in representing the value of rewards (and many other stimuli) is that the number of possible values has no absolute limits. By contrast, the representational capacity of the brain is limited, as exemplified by its finite number of neurons and the limited number of possible spike outputs of each neuron. If a neuron's limited outputs were allocated evenly to represent the large, potentially infinite number of possible reward values, then that neuron's activity would allow for little if any discrimination between rewards. However, a neuron's discriminative capacity can be improved if the neuron has access to information indicating that some reward values are more likely to occur than others and if it can allocate most of its spike outputs to representing the most probable values. Conditioned, reward-predicting stimuli could provide such information for neurons, as they do in a more general way for behavior (*1–3*). Here we investigate how dopamine neurons adapt to the information about reward value contained in predictive stimuli. These neurons play a major role in

Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK, and Institute of Physiology, University of Fribourg, CH-1700 Fribourg, Switzerland.

*Present address: Department of Biology, Stanford University, Stanford, CA 94305–5125, USA.
†To whom correspondence should be addressed. E-mail: ws234@cam.ac.uk

reward processing (*4–7*) and respond to rewards and reward-predicting stimuli (*8–11*).

We presented distinct visual stimuli that specified both the probability and magnitude of otherwise identical juice rewards to monkeys well trained in a Pavlovian procedure (*12*). Standard procedures were employed to extracellularly record the activity of single dopamine neurons of midbrain groups A8, A9, and A10 in two awake Macaque monkeys (*12*). We report data for all recorded neurons that displayed electrophysiological characteristics typical of dopamine neurons (wide impulses at low rates) (*12, 13*). In an attempt to accurately portray the whole population of dopamine neurons, we did not select neurons on the basis of their modulation by a reward event.

The expected value of future rewards (the sum of possible reward magnitudes, each weighted by its probability) is thought to be an important variable determining choice behavior (*14–17*). To test this, we trained an animal with a set of five distinct visual stimuli presented in pseudorandom alternation. Each stimulus indicated the probability that a specific liquid volume would be delivered 2 s after stimulus onset. Anticipatory licking before liquid delivery was elicited by the smallest positive expected liquid volume tested (0.05 ml at probability $p = 0.5$) and increased with expected liquid volume, suggesting that the animals had learned to use the stimuli to predict liquid delivery and that the larger liquid volumes corresponded to larger reward values (Fig. 1A). The transient

activation of dopamine neurons increased monotonically with the expected liquid volume associated with each stimulus (Fig. 1, B and C). For example, the stimulus predicting 0.15 ml at $p = 1.0$ elicited significantly greater neural activation than the stimulus predicting the same magnitude reward at $p = 0.5$, but less activation than the stimulus predicting 0.50 ml at $p = 0.5$. The activation of dopamine neurons also increased with the combination of magnitude and probability when the stimuli predicted that either of two nonzero magnitudes would occur with equal probability (Fig. 1C, animal B).

To investigate whether individual neurons might be preferentially sensitive to probability or magnitude, we took independent measures of sensitivity to magnitude and probability in each neuron ($n = 57$ neurons). There was a positive correlation ($R^2 = 0.23$, $P < 0.005$), indicating that those neurons that were most sensitive to reward magnitude were also most sensitive to probability (Fig. 1D). Thus, it appears that dopamine neurons encode a combination of magnitude and probability, as expressed, for example, by the expected reward value, rather than distinguishing between the two.

Having examined responses to reward-predicting stimuli of differing values, we investigated the extent to which dopamine neurons discriminated between different volumes of unpredicted liquid. We delivered three distinct liquid volumes (0.05, 0.15, and 0.50 ml) in pseudorandom alternation with a variable intertrial interval (*18*) and in the absence of any explicit predictive stimuli. Both individual dopamine neurons (43 of 55 neurons tested; $P < 0.01$, Wilcoxon test) and the population as a whole (55 neurons) showed greater activation for the large than for the small liquid volume (Fig. 2). Thus, the activation of dopamine neurons increased with the reward value of unpredicted liquids, similar to the responses to reward-predicting visual stimuli.

Although these results suggest that dopamine neurons encode the reward value in a monotonically increasing fashion, past work indicates that they do not represent absolute value. Rather, they appear to encode value as a prediction error by representing at each moment in time the difference between the
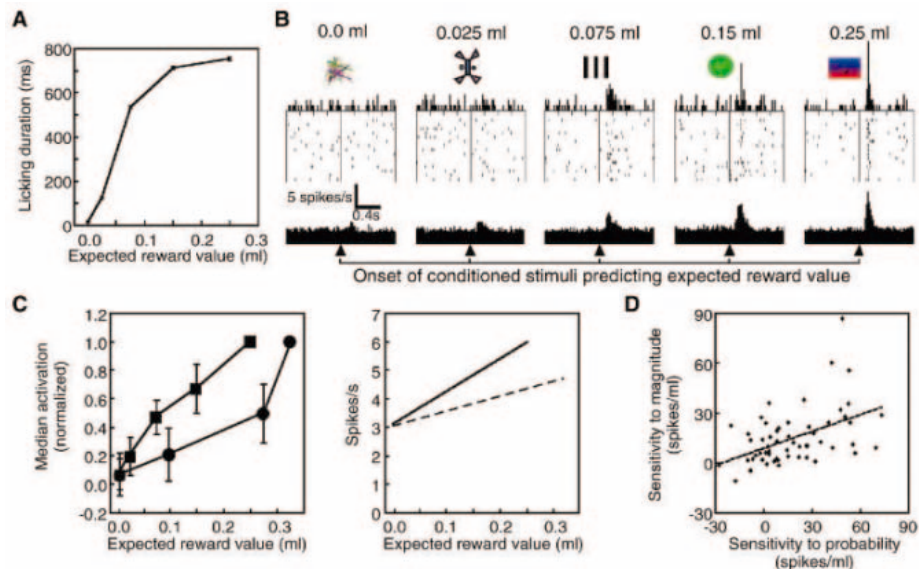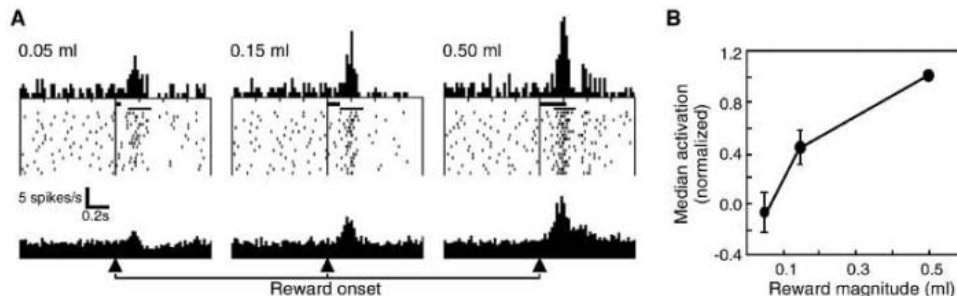
Fig. 1. Behavioral and neuronal responses to conditioned stimuli increase with expected reward value. (**A**) Anticipatory licking responses during the 2-s delay between the conditioned stimuli and liquid delivery. Each point shows the mean (± SEM) of at least 1835 trials (animal A) and is significantly different from all other points (*t* tests). Similar results were obtained from animal B, although the mean licking durations varied over a smaller range. (**B**) Single-neuron (top) and population responses (bottom) (*n* = 57 neurons) from the experiment in (A). Visual conditioned stimuli with their expected magnitude of reward are shown above the rasters. Expected values (probability × magnitude) were, from left to right, 0 ml (1.0 probability × 0.0 ml magnitude), 0.025 ml (0.5 × 0.05 ml), 0.075 ml (0.5 × 0.15 ml), 0.15 ml (1.0 × 0.15 ml), and 0.25 ml (0.5 × 0.50). Bin width is 10 ms in histograms of all figures. (**C**) (Left) Population responses as a function of expected liquid volume. Measurements were taken 90 to 180 ms (animal A) and 110 to 240 ms (animal B) after the onset of visual stimuli. The median (±95%

confidence intervals) percent change in firing rates within the population was calculated after normalization of responses within each neuron to the response evoked after onset of the stimulus associated with the largest expected value. This stimulus elicited a median activation of 167% in animal A (*n* = 57 neurons) and 40% in animal B (*n* = 53 neurons). For animal A (squares), stimuli indicated probability and magnitude as in (B). For animal B (circles), one stimulus was never followed by liquid, whereas each of the other three stimuli was associated with two volumes of equal probability (0.05 or 0.15 ml, 0.05 or 0.50 ml, and 0.15 or 0.50 ml). In each animal, the population of neurons discriminated among each expected value tested, except for 0.0 versus 0.025 ml in animal A. (Right) An alternative analysis, illustrating the sensitivity (spikes/s/ml) of a typical dopamine response to expected liquid volume. For each individual neuron, the number of impulses after stimulus onset was plotted as a function of expected magnitude, and a line was fit. The lines shown are the median lines of each population of neurons (animal A, solid line, spikes/s = 11.5 × magnitude + 3.1, R^2 = 0.51; animal B, spikes/s = 5.2 × magnitude + 3.0, R^2 = 0.69). (**D**) Positive correlation between the sensitivity of individual neurons to reward probability and magnitude (R^2 = 0.23, P < 0.005). For the data from animal A in (C), responses in each neuron (*n* = 57 neurons) are plotted both as a function of expected value, as determined both by reward probability (0.15 ml at p = 0.0, 0.5, and 1.0) and by liquid volume (0.05, 0.15, and 0.50 ml at p = 0.5). A line was fit in each case, and the slopes provided independent estimates of the sensitivity of that neuron to reward probability and magnitude. For each neuron, the slopes are plotted against each other.

Fig. 2. Neural discrimination of liquid volume. (**A**) (Top) Rasters and histograms of activity from a single dopamine neuron. (Bottom) Population histograms of activity from all neurons tested (*n* = 55 neurons). Three volumes of liquid were delivered in pseudorandom alternation in the absence of any explicit predictive stimuli. The intertrial interval ensured that the expected volume at any given moment was low (*18*). Thick horizontal bars above the rasters indicate the time of reward delivery, and thin horizontal bars indicate the single

standard time window that was used for measuring the magnitude of all responses in all neurons, as summarized in (B). Similar windows were used for all analyses and plots (supporting text). (**B**) Neural response as a function of liquid volume. Median (±95% confidence intervals) percentage change in activity for the population of neurons (*n* = 55 neurons) was calculated for responses to each volume after normalization in each neuron to the response after delivery of 0.5 ml, which itself elicited a median activation of 159% above baseline activity.

reward value (the sum of current and future rewards) and its expected value (before observation of current sensory input). Recent work demonstrates that, when signaling prediction errors, dopamine neurons are able to use contextual information in addition to information from explicitly conditioned stimuli (*19*). In the experiments shown in Figs. 1 and 2, all visual stimuli and liquid volumes were delivered in a context in which the expected reward value at each moment in time was low and invariant across trial types because of the intertrial interval (*18*). In our next set of experiments, we delivered differ-

ent volumes of liquid in the presence of explicit predictions indicated by conditioned stimuli, allowing us to systematically vary the expected value and range of reward.

Consistent with past work, a reward occurring exactly at the expected value (0.15 ml) elicited no response. However, when liquid volume was unpredictably smaller (0.05 ml) or larger (0.50 ml) in a minority of trials, dopamine neurons were suppressed or activated, respectively, compared to both the prestimulus baseline and the response to the expected volume delivered in the majority of trials (P < 0.01, Mann-Whitney test) (Fig. 3,

A and B). In an additional experiment, one stimulus predicted that either the small or medium volume would be delivered with equal probability, whereas another stimulus predicted either the medium or large volume with equal probability. In both cases, delivery of the larger of the two potential volumes elicited an increase in activity, whereas the smaller volume elicited a decrease (Fig. 3C). Thus, the identical medium volume had opposite effects on activity depending on the prediction (P < 0.01 in 19 of 53 neurons, Mann-Whitney; P < 0.0001 for the population of 53 neurons, Wilcoxon test) (Fig. 3D). These

results show how dopamine neurons process reward magnitude relative to a predicted magnitude and that a reward outcome that is positive on an absolute scale can nonetheless suppress the activity of dopamine neurons.
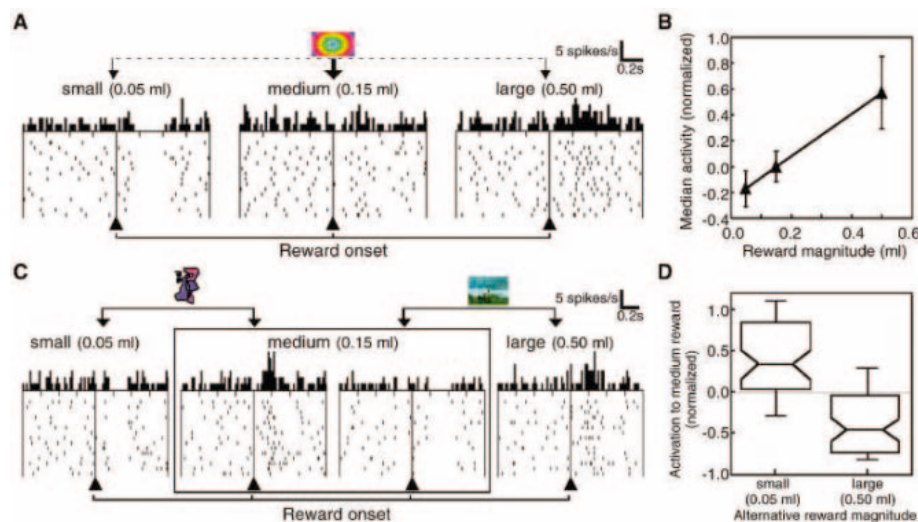
Although these results suggest that dopamine responses shift relative to the predicted reward magnitude, it is not known how their activity scales with the difference between actual and expected reward. To this end, we analyzed the dopamine responses at the time of the reward in the experiment shown in Fig. 1. Each of three distinct visual stimuli, presented on pseudorandomly alternating trials, predicted that one of two potential liquid volumes would be delivered with equal probability. Animals discriminated behaviorally between the three reward-predicting stimuli (Fig. 1A). Confirming the data described above, the larger of the two volumes always elicited an increase in activity at the time of the reward, and the smaller a decrease. However, the magnitude of activation or suppression appeared to be identical in each case, despite the fact that the absolute difference between actual and expected volume varied over a 10-fold range (Fig. 4, A and B). Thus, the responses of dopamine neurons did not appear to scale according to the absolute dif-

ference between actual and expected reward. Rather, the sensitivity or gain of the neural responses appeared to adapt according to the discrepancy in volume between the two potential outcomes.
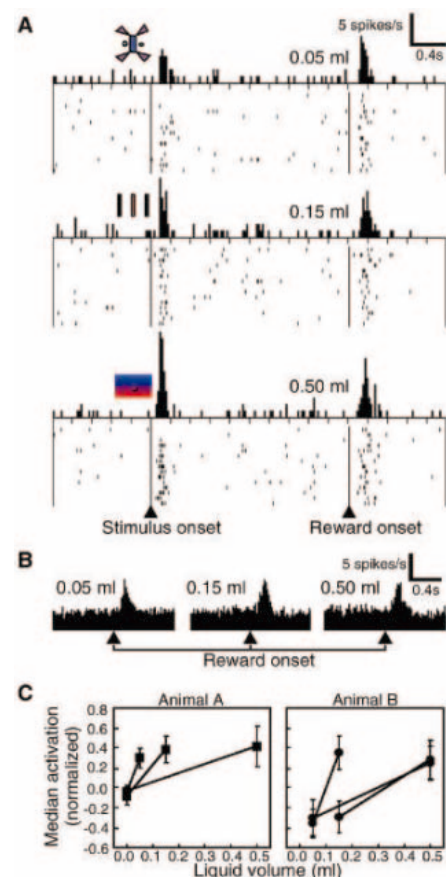
To document this result further, we plotted the median neural responses as a function of liquid volume and drew a straight line to connect the data points representing the larger and smaller outcomes after each visual stimulus (Fig. 4C). The slope of these lines provided an estimate of the neurons' gain or sensitivity with respect to liquid volume. When the discrepancy was large, the sensitivity of dopamine neurons was low, and when the discrepancy was small, sensitivity was high. As a result of this adaptation, the neural responses discriminated between the two likely outcomes equally well, regardless of their absolute difference in magnitude. The present data are not sufficient to determine precisely to which aspect of the reward prediction the neuron's sensitivity adapted, but further analysis provided limited evidence that sensitivity adapted to the discrepancy between potential liquid volumes (such as the difference or variance) rather than to their expected value (12) (fig. S2).

Our results suggest that the activity of dopamine neurons carries information on the

magnitude of reward. In representing reward magnitude, neural activity displayed two forms of adaptation that depended on the prediction that was in place at the time of the reward. First, the activity increased or de-

Fig. 4. Neural sensitivity to liquid volume adapts in response to predictive stimuli. (**A**) Activity of a single neuron showing nearly identical responses to three liquid volumes spanning a 10-fold range. Each of three pseudorandomly alternating visual stimuli (shown at left) was followed by one of two liquid volumes at $p = 0.5$ (top, 0.0 or 0.05 ml; middle, 0.0 or 0.15 ml; bottom, 0.0 or 0.5 ml). Responses after onset of visual stimuli increased with their associated expected reward values. Only rewarded trials are shown. (**B**) Population histograms for different liquid volumes from the experiment in (A) (57 neurons, animal A). (**C**) Each line connects responses occurring in the context of a specific conditioned stimulus, and its slope provides a measure of gain or sensitivity. Each point represents the median (±95% confidence intervals) response of the population taken after normalizing the percentage change in activity in each neuron to the response after unpredicted liquid (0.15 ml) delivered in a separate block of trials (which elicited an activation of 266% above baseline in animal A, $n = 57$ neurons, and 97% in animal B, $n = 53$ neurons). (Left) The experiment in (A) and (B). (Right) The same experiment, but performed in animal B with two nonzero liquid volumes per conditioned stimulus at equal probability ($p = 0.5$) (stimulus 1: 0.05 versus 0.15 ml, stimulus 2: 0.15 versus 0.5 ml, stimulus 3: 0.05 versus 0.5 ml).

Fig. 3. Bidirectional dopamine responses to reward outcomes reflect deviations from predictions. (**A**) A single conditioned stimulus was usually followed by an intermediate volume of liquid (0.15 ml) that elicited no change in the neuron's activity (center). However, on a small minority of trials, smaller (0.05 ml) or larger (0.50 ml) volumes were unpredictably substituted, and neural activity decreased (left) or increased (right), respectively. Neural responses to the large liquid volume were relatively long-lasting (supporting online text). (**B**) Median responses (±95% confidence intervals) from the population as a function of liquid volume for the experiment in (A) (12 neurons from animal A, 17 neurons from animal B). Responses in each neuron were normalized to the response after the unpredicted delivery of liquid (0.15 ml) in a separate block of trials and in the absence of any explicit reward-predicting stimulus. (**C**) Responses of a single neuron to three liquid volumes, delivered in the context of two different predictions. One stimulus predicted small or medium volume with equal probability, whereas another stimulus predicted medium or large volume. The medium volume activated the neuron in one context, but suppressed activity in the other. (**D**) Population responses ($n = 53$ neurons, animal B) to medium reward in the experiment in (C). The plot shows the median, the ±95% confidence intervals (notches corresponding to obtuse angles), the 25th and 75th percentiles (boundaries corresponding to right angles), and the 10th and 90th percentiles (bars). In each neuron, percentage change in activity was normalized to the response to unpredicted liquid (0.15 ml, which elicited a median increase in activity of 97%).

creased depending on whether the reward outcome was larger or smaller, respectively, than an intermediate reference point such as expected value. A second, unanticipated form of adaptation was the change in sensitivity or gain of neural activity that appeared to depend on the range of likely reward magnitudes (Fig. 4). Thus, the larger of two potential rewards always elicited the same increase in activity and the smaller of the two elicited the same decrease in activity, regardless of absolute magnitude. The identical responses to liquid volumes spanning a 10-fold range were not due to an insensitivity of the dopamine neurons, which were capable of greater activations (Fig. 4C, note normalization of data points) and discriminated well among these same liquid volumes when delivered in the absence of explicit predictive stimuli (Fig. 2). Rather, the gain of neural activity with respect to liquid volume appeared to adapt in proportion to the range or standard deviation of the predicted reward outcomes, so that neural discrimination between the two reward outcomes that were most probable from the animal's perspective was robust regardless of their absolute difference in magnitude.

The efficiency and accuracy with which neural activity can code the value of a stimulus (such as liquid volume) can be greatly increased if neurons make use of information about the probabilities of potential reward values. Neural activity can then be devoted to representing probable values at the expense of improbable values. Our evidence suggests that the transient dopamine response to conditioned stimuli may carry information on expected reward value, and previous work shows that the more sustained activity of dopamine neurons reflects a measure of reward uncertainty such as variance (10). If the system possesses prior information consisting of the expected value and variance of reward, then this information need not be represented redundantly at the time of re-ward. Discarding this old information may be achieved by subtracting the expected value from the absolute reward value and then dividing by the variance. Analogous normalization processes appear to occur in early visual neurons (20–22). It is not known to what extent the normalization processes observed in dopamine neurons are actually performed in dopamine neurons as opposed to their afferent input structures (23). Because the new information is by definition precisely the information that the system needs to learn, the activity of dopamine neurons would be an appropriate teaching signal (24).

Adaptation appears to be a nearly universal feature of neural activity. There is substantial evidence, particularly from the early visual system, that adaptation contributes to the efficient representation of stimuli (20–22, 25–28). We have extended the principles of efficient representation to the study of reward. Reward is central to processes underlying behavior, such as reinforcement learning and decision-making, and consideration of limitations and efficiency in the neural representation of reward may yield insights into these processes.

References and Notes
1. I. P. Pavlov, *Conditional Reflexes* (Oxford Univ. Press, Oxford, 1927).
2. E. L. Thorndike, *Animal Intelligence: Experimental Studies* (Macmillan, New York, 1911).
3. R. A. Rescorla, A. R. Wagner, in *Classical Conditioning II: Current Research and Theory*, A. Black, W. F. Prokasy, Eds. (Appleton-Century-Crofts, New York, 1972), pp. 64–99.
4. H. C. Fibiger, A. G. Phillips, in *The Nervous System*, vol. 4 of *Handbook of Physiology*, F. E. Bloom, Ed. (Williams & Wilkins, Baltimore, MD, 1986), pp. 647–675.
5. R. A. Wise, D. C. Hoffman, *Synapse* **10**, 247 (1992).
6. T. E. Robinson, K. C. Berridge, *Brain Res. Rev.* **18**, 247 (1993).
7. T. W. Robbins, B. J. Everitt, *Curr. Opin. Neurobiol.* **6**, 228 (1996).
8. P. Waelti, A. Dickinson, W. Schultz, *Nature* **412**, 43 (2001).
9. T. Satoh, S. Nakai, T. Sato, M. Kimura, *J. Neurosci.* **23**, 9913 (2003).
10. C. D. Fiorillo, P. N. Tobler, W. Schultz, *Science* **299**, 1898 (2003).
11. R. Kawagoe, Y. Takikawa, O. Hikosaka, *J. Neurophysiol.* **91**, 1013 (2004).
12. Materials and methods are available as supporting material on *Science* Online.
13. No significant correlations were found between neuronal position in areas A8, A9, and A10 and the different types of responses in all the experiments reported; thus, the data were pooled.
14. A. Arnauld, P. Nicole, *La logique, ou, L'art de penser* (Librairie philosophique J. Vrin, Paris, 1981/1662).
15. P. W. Glimcher, *Decisions, Uncertainty and the Brain* (MIT Press, Cambridge, MA, 2003).
16. P. Shizgal, *Curr. Opin. Neurobiol.* **7**, 198 (1997).
17. M. I. Leon, C. R. Gallistel, *J. Exp. Psychol. Anim. Behav. Process.* **24**, 265 (1998).
18. All trials were separated by an average intertrial interval of 9 s, consisting of a fixed 4 s plus an interval drawn from a truncated Poisson-like distribution with a mean of 5 s. Thus, the probability that a trial would begin at any given moment was low.
19. H. Nakahara, H. Itoh, R. Kawagoe, Y. Takikawa, O. Hikosaka, *Neuron* **41**, 269 (2004).
20. M. V. Srinivasan, S. B. Laughlin, A. Dubs, *Proc. R. Soc. London Ser. B.* **216**, 427 (1982).
21. N. Brenner, W. Bialek, R. R. de Ruyter van Steveninck, *Neuron* **26**, 695 (2000).
22. A. L. Fairhall, G. D. Lewen, W. Bialek, R. R. de Ruyter van Steveninck, *Nature* **412**, 787 (2001).
23. W. Schultz, *J. Neurophysiol.* **80**, 1 (1998).
24. W. Schultz, P. Dayan, R. R. Montague, *Science* **275**, 1593 (1997).
25. H. B. Barlow, in *Sensory Communication*, W. A. Rosenblith, Ed. (MIT Press, Cambridge, MA, 1961), pp. 217–234.
26. S. B. Laughlin, *Z. Naturforsch. Teil C Biochem. Biophys. Biol. Virol.* **36**, 910 (1981).
27. I. Ohzawa, G. Sclar, R. D. Freeman, *Nature* **298**, 266 (1982).
28. S. M. Smirnakis, M. J. Berry, D. K. Warland, W. Bialek, M. Meister, *Nature* **386**, 69 (1997).
29. We thank K. I. Tsutsui, I. Hernadi, A. Dickinson, and S. B. Laughlin for helpful comments. Supported by the Wellcome Trust, Swiss National Science Foundation (W.S. and P.N.T.), Roche Research Foundation (P.N.T.), Janggen-Poehn Foundation (P.N.T.), and Human Frontiers Science Program (C.D.F.).

Protein Purification for Small-Scale Research

The AKTAxpress Twin, the second in a series of fully integrated protein purification systems, is designed to deliver totally automated, multi-step purification of histidine- and glutathione-S-transferase–tagged proteins for small-scale research. It is a smaller version of the AKTAxpress, which is a high-throughput protein purification system for large-scale drug discovery processes. The new system is designed to deliver the same benefits to researchers who do not require the high-throughput capacity. It includes a software wizard that selects the optimum purification protocol and automatically transfers the largest peak from every purification step to the next column. This multi-step process provides a protein purity greater than 95%. It is a two-module system that accepts multiple samples, offering researchers the ability to purify eight samples in about 11 hours.

GE Healthcare For information 732-457-8149 www.gehealthcare.com

UV-Visible Spectrophotometer

The Libra S35PC Ultraviolet-Visible Spectrophotometer has a high-performance, 1-nm bandwidth optical system. Optical specifications exceed Pharmacopoeia requirements over the operating range of 190 nm to 1100 nm. The instrument offers built-in performance validation. Other hardware features include unique "Press to Read" deuterium lamp technology, an innovation that increases source lifetime by only running the source at maximum energy during an actual measurement sequence, and an eight-position sample changer.

Biochrom For information +44 (0) 1223 423723 www.biochrom.co.uk

Manual Stereotaxic Injector

The Manual Stereotaxic Injector provides a simple, economical, and effective method for injecting or withdrawing calibrated volumes. The manual control allows researchers to precisely control the infusion of volumes as small as 200 picoliters. It includes a large probe holder for mounting easily on most stereotaxic instruments. It injects directly through the needle on the syringe, which prevents degradation and avoids compliance issues that arise when injections are forced through plastic tubing.

Stoelting For information 630-860-9700 www.stoeltingco.com

EliSpot Slide

The EliSpot Slide offers a convenient, microscope slide format for measuring immune cell activation at the single-cell level. It features a removable 16-well chamber and lid made of low-protein-binding material that form a leak-free barrier between wells. The standard slide design facilitates quick and efficient washing, handling, and sample storage. A glass surface provides optimum flatness for scanning results.

Grace Bio-Labs For information 800-813-7339 www.gracebio.com

Benchtop Shell Bath

The VirTis Benchtop Shell Bath is a self-contained, compact freezing unit that produces a frozen product coating to the inside of freeze-drying flasks. This allows the freeze-drying process to be more efficient than in conditions such as "stub freezing." A thin coating of product is evenly frozen around the inside of a freeze-drying flask, permitting a greater volume of material to be quickly and efficiently processed during each freeze-drying run. The freeze-dry cycle is shortened because moisture does not have to migrate through the ice interface and a thick layer of already-dried product. The higher surface area also allows better heat transfer.

SP Industries For information 800-523-2327 www.SPindustries.com

High-Flow Plastic Filters

A new series of black or white polypropylene filters is available. Suitable for air, liquid, or vacuum service, these quickly disassembled filters are easily serviceable by replacing the filter materials. Filters are offered in filtration of 15, 30, 70, 110, and 170 µm porous disk material. Nylon screen filters are also available. Pressures to 125 psi and temperatures to +140°F can be attained.

Industrial Specialties Mfg For information 303-781-8486 www.industrialspec.com

High-Throughput GC Injection System

The AOC-5000 High Throughput Gas Chromatography (GC) Injection System combines static headspace, liquid injection, and solid-phase micro-extraction modes in one instrument. This unique combination allows quick switching from one application to another on the same GC system, enhancing the speed, flexibility, capacity, and precision of the instrument. Top-mounted on GCs to save bench space, the system is designed with a syringe-only concept for transparent sample operation with sample loops, transfer lines, or switching valves.

Shimadzu For information 800-477-1227 www.ssi.shimadzu.com

Liquid Handling Workstations

Three new epMotion 5075 robotic liquid handling workstations feature an innovative optical sensor to monitor labware and liquid levels on the 12-place deck, the ability to automatically change pipetting heads and move labware to various positions, and a choice of control panel or personal computer control operation. Base Model LH can be upgraded later to Model VAC, which has a built-in manifold and vacuum pump for walk-away automation of nucleic acid purification, or to Model MC, which has an integrated MasterCycler ep 96 or 384 for complete automation of polymerase chain reaction and sequence reactions.

Eppendorf For information 800-645-3050 www.brinkmann.com

For more information visit **GetInfo**, *Science*'s new online product index at *http://science.labvelocity.com*

From the pages of GetInfo, you can:
- Quickly find and request free information on products and services found in the pages of *Science*.
- Ask vendors to contact you with more information.
- Link directly to vendors' Web sites.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting www.science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.